

Access DB#

71413

## SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: NA WALICKA Examiner #: 78201 Date: July 19, 2002  
Art Unit: 1652 Phone Number 305-7270 Serial Number: 04/657,986  
Mail Box and Bldg/Room Location: 10000 Results Format Preferred (circle): PAPER DISK E-MAIL  
10001

If more than one search is submitted, please prioritize searches in order of need.

\*\*\*\*\*

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Inhibitors of SERINE PROTEASE ACTIVITY

Inventors (please provide full names): Madison et al

Earliest Priority Filing Date: 09/08/2000

\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Please search

SEQ ID NO: 2, ASAP please

Thank you.

Intended

(S) (U)

RECEIVED  
JUL 19 2002

Point of Contact:  
Toby Port  
Technical Info. Specialist  
CM1 6A04  
703-308-3534

## STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher: _____	NA Sequence (#) _____	STN _____
Searcher Phone #: _____	AA Sequence (#) <u>1</u>	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: <u>7/22</u>	Bibliographic _____	Dr. Link _____
Date Completed: <u>7/23</u>	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: <u>10</u>	Fulltext _____	Sequence Systems <u>CS</u>
Clerical Prep Time: _____	Patent Family _____	WWW/Internet _____
Online Time: <u>11</u>	Other _____	Other (specify) _____

GenCore version 4.5.  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2002, 17:07:54 ; Search time 29.61 seconds  
(without alignments)  
904.045 Million cell updates/sec

Title: US-09-657-986b-2

Perfect score: 1319

Sequence: 1 VVGSTDADEGEWPCVSLHA.....PGVYTRLPFRDWIKENTGV 241

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_032802:\*  
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1319	100.0	241	AAE06936	Human membrane-type
2	1319	100.0	683	AA19551	Human matriptase
3	1319	100.0	762	AA190284	Human peptidase, H
4	1319	100.0	851	AB111428	Human membrane-type
5	1319	100.0	851	AA25628	Human protein sequ
6	1319	100.0	855	AA106671	Tumour antigen der
7	1319	100.0	855	AA19552	Human matriptase
8	1319	100.0	855	AAE06930	Human membrane-type
9	1319	100.0	855	AA198500	Human TADG-15, HO
10	1319	100.0	855	AA135465	Human membrane-type
11	1315	99.7	241	AAW22987	Human serine prote

12	1227	93.0	932	22	ABG21442	Novel human diagno
13	1100	83.4	902	22	AA198507	Marine epithin. M
14	643	48.7	129	22	ABG21439	Novel human diagno
15	576.5	43.7	233	22	AAE01942	Human endotheliase
16	576.5	43.7	422	21	AA194708	Human DESCI protei
17	576.5	43.7	423	21	AA199414	Human PRO1461 (UNQ
18	576.5	43.7	423	22	AAU29183	Human PRO polysept
19	576.5	43.7	423	22	AAU01344	Human TANGO 361 am
20	576.5	43.7	423	22	AAU01399	Human TANGO 361, v
21	576.5	43.7	423	22	AAU01400	Human TANGO 361, v
22	576.5	43.7	423	22	AAU01401	Human TANGO 361, v
23	576.5	43.7	423	22	AAU01402	Human TANGO 361, v
24	576.5	43.7	423	22	AA187578	Human PRO1461. HO
25	576.5	43.7	423	22	AA186613	Protein of the inv
26	563.5	42.6	422	22	AAE01946	Human endotheliase
27	562.5	42.6	422	21	AA194709	Human DESCI protei
28	561.5	42.6	235	22	AAE06932	Human membrane-type
29	561.5	42.6	658	22	AAE06934	Human membrane-type
30	561.5	42.6	802	22	AAE06933	Human membrane-type
31	559.5	42.4	802	20	AA141710	Human PRO618 prote
32	559.5	42.4	802	21	AA144266	Human PRO618 (UNQ3
33	559.5	42.4	802	21	AA124052	Human PRO618 prote
34	547.5	41.5	1019	22	AAE06940	Human enterokinase
35	544.5	41.3	798	15	AA157283	Bovine enterokinas
36	540	40.9	238	22	AAE06937	Novel human diagno
37	506	38.4	145	21	AA190298	Human peptidase, H
38	501.5	38.0	250	22	AB165923	Drosophila melanog
39	499	37.8	492	21	AA136901	Human TMPRSS2 prot
40	499	37.8	492	21	AA157280	Ovrl15 homolog pro
41	499	37.8	492	22	AA169939	Human transmembran
42	499	37.8	492	22	AAE06943	Human serine prote
43	499	37.8	492	22	AA101294	Human transmembran
44	497.5	37.7	232	17	AA189430	Trypsin-like enzym
45	497.5	37.7	418	17	AA189435	Trypsin-like enzym

ALIGNMENTS

RESULT 1  
AAE06936  
ID AAE06936 standard; Protein; 241 AA.  
AC AAE06936;

DT 16-OCT-2001 (first entry)

DE Human membrane-type serine protease (WTSP) 1 protease domain.

DE Human; transmembrane serine protease; membrane-type serine protease;  
KW WTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic;  
KW lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy;  
KW matriptase.

XX Homo sapiens.

OS Homo sapiens.

XX WO200157194-A2.

PN WO200157194-A2.

PD 09-AUG-2001.

XX 02-FEB-2001; 2001WO-US03471.

XX 03-FEB-2000; 2000US-0179982.

PR 18-FEB-2000; 2000US-0183542.

PR 22-JUN-2000; 2000US-0213124.

PR 26-JUL-2000; 2000US-0220970.

PR 08-SEP-2000; 2000US-0657986.

PR 22-SEP-2000; 2000US-0234640.

XX (CORV-) CORVAS INT INC.

PA Madison EL, Ong EO, Yeh J;

PI

```

DR WPI; 2001-488877/53.
DR N-PSDB; AAD13155.
XX
XX Novel single chain polypeptide comprising protease domain of type-II
XX membrane-type serine protease or its catalytically active portion
XX useful for treating and preventing cancer and tumor
XX
PS Example 5; Page 227; 256pp; English.
XX
XX The invention relates to transmembrane serine proteases and their
XX corresponding nucleotides and the protease domain of a type-II
XX membrane-type serine protease (MTSP). MTSP is useful for identifying
XX compounds that modulate or inhibit its proteolytic activity and for
XX formulating a medicament for treating neoplastic disease. MTSP and
XX its corresponding nucleotides are useful in preventing or treating
XX tumours or cancers such as lung carcinoma, colon adenocarcinoma and
XX ovarian carcinoma, in diagnostics and in hybridisation assays. MTSP
XX is useful as a diagnostic marker for tumour development, growth and/or
XX progression and as immunogens to generate antibodies that specifically
XX bind to it. MTSP DNA is useful in a yeast two-hybrid system and in
XX gene therapy. The present sequence is protease domain of human MTSP1
XX protein (also called matriptase).
SQ Sequence 241 AA;

Query Match 100.0%; Score 1319; DB 22; Length 241;
Best Local Similarity 100.0%; Pred. No. 1.1e-98;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVGTDADGEQWVQVSLHALGQGHICGASLISPNWLVSAAHCHYIDRGFRYSDPTQWTA 60
Db 1 VVGTDADGEQWVQVSLHALGQGHICGASLISPNWLVSAAHCHYIDRGFRYSDPTQWTA 60
QY 61 FLGLHDSQSASQVQERLKRITSHPTNDFDYDIALLEKPAEYSSMVAPICLPD 120
Db 61 FLGLHDSQSASQVQERLKRITSHPTNDFDYDIALLEKPAEYSSMVAPICLPD 120
QY 121 ASHVFPAGKAIWVGHTQYGGTGALILQKEIRVINOTTCEMLPQQTIPRMVCGFL 180
Db 121 ASHVFPAGKAIWVGHTQYGGTGALILQKEIRVINOTTCEMLPQQTIPRMVCGFL 180
QY 181 SGGVDSQSGSGPLSSVEADGRIFQAGVSVWGDCACORNGPVYVTRLPFRWIKENTG 240
Db 181 SGGVDSQSGSGPLSSVEADGRIFQAGVSVWGDCACORNGPVYVTRLPFRWIKENTG 240
QY 241 V 241
Db 241 V 241

RESULT 2
AAB19551
ID AAB19551 standard; Protein; 683 AA.
XX
XX AAB19551;
XX
XX 22-JAN-2001 (first entry)
XX
XX Human matriptase (truncated form).
XX
XX Matriptase; serine protease; human; breast cancer; pre-malignancy;
XX actinic keratosis; leukoplakia; Barrett's epithelium;
XX colunar metaplasia; ulcerative colitis; Bowenoid papulosis;
XX adenomatous colorectal polyp; Ovarat erythroplasia;
XX vulvar intraepithelial neoplasia; tumour; metastasis; therapy.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Peptide 1..31
XX Protein 32..683
  
```

---

```

FT Region /label= Mature_protein
FT 42..155 /note= "complement subcomponent 1r and 1s (Clr/s)
FT region"
FT 168..268 /note= "complement subcomponent 1r and 1s (Clr/s)
FT region"
FT 280..314 /note= "LDL-receptor type region"
FT 315..351 /note= "LDL-receptor type region"
FT 352..387 /note= "LDL-receptor type region"
FT 394..430 /note= "LDL-receptor type region"
FT 432..683 /label= Catalytic_domain
FT Modified-site /note= "N-glycosylated"
FT Modified-site /note= "N-glycosylated"
FT Active-site /note= "forms catalytic triad with Asp-539 and
FT Ser-633"
FT Active-site /note= "forms catalytic triad with His-484 and
FT Ser-633"
FT Active-site /note= "forms catalytic triad with His-484 and
FT Asp-539"
FT Peptide 77..79 /note= "RGD motif"
FT Disulfide-bond 469..485
FT Disulfide-bond 604..618
FT Disulfide-bond 629..658
FT Cleavage-site 442..446 /note= "conserved proteolytic activation site"
XX
XX W0200053232-A1.
XX
XX 14-SEP-2000.
XX
XX 10-MAR-2000; 2000WO-US06111.
XX
XX 12-MAR-1999; 99US-0124006.
XX
XX (GEOU ) UNIV GEORGETOWN.
XX
XX Dickson RB, Lin C, Johnson M, Wang S, Enyedy I;
XX
XX WPI; 2000-594268/56.
XX N-PSDB; AAA88492.
XX
XX Treating malignancies, premalignant and pathologic conditions in a
XX subject, comprises administering matriptase modulating agent
XX
XX Claim 14; Fig 9; 116pp; English.
XX
XX The present sequence is that of the truncated form of human
XX matriptase, a trypsin-like protease, as deduced from cDNA (see
XX AAA88492) obtained from human breast cancer cell cDNA by RT-PCR.
XX The full-length form (see AAB19552) has an additional 172 N-terminal
XX amino acids. Either form can be produced in transformed or
XX transfected cells using a claimed method. The zymogen (inactive)
XX form of matriptase is a single-chain protein. The active 2-chain
XX form strongly interacts with fragments of a Kunitz-type serine
XX protease inhibitor (hepatocyte growth factor activator inhibitor,
XX HAI-1) to form SDS-stable complexes. In breast cancer cells,
XX matriptase is present mainly as the uncomplexed form. Only the
XX complexed matriptase is detected in human milk. The invention is
XX directed to a method of detecting a malignancy or a pre-malignant
XX lesion in breast or other tissue by detecting the presence of
XX single- or 2-chain forms of matriptase in the tissue. The
  
```

CC object is to inhibit tumour onset, tumour growth and metastasis.  
 CC Malignancies and pre-malignant conditions characterised by  
 CC expression of the zymogen or activated form of matrilysin  
 CC created by administering an inhibitor of matrilysin, especially a  
 CC Bowman-Birk inhibitor. The pre-malignant condition is atypical  
 CC ductal hyperplasia of the breast, actinic keratosis, leukoplakia,  
 CC Barrett's epithelium of the esophagus, ulcerative colitis,  
 CC adenomatous colorectal polyps, erythroplasia of the Queyrat,  
 CC Bowen's disease, Bowenoid papulosis, vulvar intraepithelial  
 CC neoplasia or dysplastic changes to the cervix. The invention also  
 CC provides methods for in vivo or in vitro diagnosis of malignancy  
 CC or pre-malignant lesion, and methods of identifying matrilysin  
 CC modulators, including activators and inhibitors.

XX Sequence 683 AA;

Query Match 100.0%; Score 1319; DB 21; Length 683;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-98;  
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VVGTDADGEWPMQVSLHALGQGHICGASLISPNMLVSAHCYIDDRGFYSDPTQWTA 60  
 Db 443 VVGTDADGEWPMQVSLHALGQGHICGASLISPNMLVSAHCYIDDRGFYSDPTQWTA 502  
 Qy 61 FLGLHDSQSRAPGVQERRLKRISHPFFNDFTDIALLELEKPAEYSSMVRPCLPD 120  
 Db 503 FLGLHDSQSRAPGVQERRLKRISHPFFNDFTDIALLELEKPAEYSSMVRPCLPD 562  
 Qy 121 ASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINTTCENLLPQGITPRMVCVGL 180  
 Db 563 ASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINTTCENLLPQGITPRMVCVGL 622  
 Qy 181 SGGVDSQSGSGPLSSVEADGRIFQAGVYVSWDGGCAQRNKPQVYTRLPFRDWIKENTG 240  
 Db 623 SGGVDSQSGSGPLSSVEADGRIFQAGVYVSWDGGCAQRNKPQVYTRLPFRDWIKENTG 682  
 Qy 241 V 241  
 Db 683 V 683

RESULT 3  
 AAY90284

ID AAY90284 standard; Protein; 762 AA.

XX AC AAY90284;

XX DT 24-OCT-2000 (first entry)

XX DE Human peptidase, HPEP-1 protein sequence.

XX KW Human; peptidase; cell proliferative disorder; arteriosclerosis;  
 KW psoriasis; myelofibrosis; cancer; autoimmune disorder; Crohn's disease;  
 KW inflammatory disorder; AIDS; anaemia; allergy; asthma; atherosclerosis;  
 KW Grave's disease; multiple sclerosis; scleroderma; infection; diabetes;  
 KW metabolic disorder; Addison's disease; cystic fibrosis; diagnosis;  
 KW glycogen storage disease; obesity; therapy; HPEP-1.

XX OS Homo sapiens.

XX PN WO200042201-A2.

XX PD 20-JUL-2000.

XX PF 11-JAN-2000; 2000WO-US00641.

XX PR 11-JAN-1999; 99US-0172247.

XX PR 03-MAY-1999; 99US-0132253.

XX PR 27-MAY-1999; 99US-0136653.

XX PA (INCY-) INCYTE PHARM INC.

PI Bandman O, Hillman JL, Tang YT, Azimzai Y, Baughn MR, Lal P;  
 PI Yue H, Lu DAM;  
 XX WPI; 2000-482832/42.  
 DR N-PSDB; AAA37657.  
 XX An isolated polypeptide for diagnosis, prevention and treatment of  
 PT cell proliferative, autoimmune/ inflammatory and metabolic disorders  
 PT comprises a sequence encoding a human peptidase -  
 XX Claim 2; Page 91-93; 131pp; English.

CC This sequence represents a human peptidase, designated HPEP-1. The  
 CC invention relates to 18 human peptidases designated HPEP-1 to HPEP-18,  
 CC respectively. The peptidases can be used for treating a disease or  
 CC condition associated with decreased expression or over expression of  
 CC functional human peptidases. The diseases that can be diagnosed,  
 CC prevented and treated include cell proliferative disorders (such as  
 CC arteriosclerosis, psoriasis, myelofibrosis, and cancers),  
 CC autoimmune/inflammatory disorders (such as AIDS, anaemia, allergies,  
 CC Crohn's disease, asthma, atherosclerosis, Grave's disease, multiple  
 CC sclerosis, and scleroderma), infections, and metabolic disorders (such as  
 CC Addison's disease, diabetes, cystic fibrosis, glycogen storage diseases  
 CC and obesity).

XX Sequence 762 AA;

Query Match 100.0%; Score 1319; DB 21; Length 762;  
 Best Local Similarity 100.0%; Pred. No. 3.4e-98;  
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VVGTDADGEWPMQVSLHALGQGHICGASLISPNMLVSAHCYIDDRGFYSDPTQWTA 60  
 Db 522 VVGTDADGEWPMQVSLHALGQGHICGASLISPNMLVSAHCYIDDRGFYSDPTQWTA 581  
 Qy 61 FLGLHDSQSRAPGVQERRLKRISHPFFNDFTDIALLELEKPAEYSSMVRPCLPD 120  
 Db 582 FLGLHDSQSRAPGVQERRLKRISHPFFNDFTDIALLELEKPAEYSSMVRPCLPD 641  
 Qy 121 ASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINTTCENLLPQGITPRMVCVGL 180  
 Db 642 ASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINTTCENLLPQGITPRMVCVGL 701  
 Qy 181 SGGVDSQSGSGPLSSVEADGRIFQAGVYVSWDGGCAQRNKPQVYTRLPFRDWIKENTG 240  
 Db 702 SGGVDSQSGSGPLSSVEADGRIFQAGVYVSWDGGCAQRNKPQVYTRLPFRDWIKENTG 761  
 Qy 241 V 241  
 Db 762 V 762

RESULT 4

ID ABB11428 standard; peptide; 851 AA.

XX AC ABB11428;

XX DT 11-JAN-2002 (first entry)

XX DE Human membrane-type Ser kinase homologue, SEQ ID NO:1798.

XX KW Human; cytokine; cell proliferation; cell differentiation; growth factor;  
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;  
 KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
 KW chronic inflammatory condition; proliferative retinopathy;  
 KW atherosclerosis; coronary heart disease; arterial ischaemia;  
 KW bone disorder; osteoporosis; vascular growth disorder;  
 KW tissue regeneration; wound healing; infection; immune disorder;  
 KW cell culture; drug screening; gene therapy; antiinflammatory;

KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;  
 KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
 KW antifungal; vulnerary; antiulcer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157188-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 05-FEB-2001; 2001WO-US03800.  
 XX  
 PR 03-FEB-2000; 2000US-0496914.  
 PR 27-APR-2000; 2000US-0560875.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 XX Tang YT, Liu C, Drmanac RT;  
 XX WPI; 2001-457740/49.  
 DR N-PSDB; ABA08672.  
 XX  
 XX Human proteins and DNA encoding sequences useful for preventing,  
 PT treating or ameliorating a medical condition in a mammalian subject  
 PT e.g. arthritis and cancer -  
 XX  
 PS Claim 20; Page 188; 1963pp; English.  
 XX  
 CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
 CC invention also relates to vectors and recombinant host cells comprising a  
 CC nucleotide of the invention, methods of producing the novel polypeptides,  
 CC antibodies against the polypeptides, methods of detecting the nucleotides  
 CC or polypeptides in a sample, and methods of identifying compounds which  
 CC bind to polypeptides of the invention. Although novel, many of the  
 CC polypeptides of the invention have homology to known proteins, thereby  
 CC giving an insight into their probable biological activities, and hence  
 CC potential therapeutic applications. The polypeptides of the invention may  
 CC have various activities, including cytokine, cell proliferation or cell  
 CC differentiation activities; stem cell growth factor activity;  
 CC haematopoiesis regulatory activity; tissue growth activity;  
 CC immunomodulatory activity; activin- or inhibin-related activities;  
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 CC thrombolytic activities; receptor or ligand activities; or may be  
 CC involved in oncogenesis, cancer cell proliferation or metastasis.  
 CC Depending on their biological activities, polypeptides and nucleotides of  
 CC the invention are useful for preventing, treating or ameliorating medical  
 CC conditions, e.g., by protein or gene therapy. Such conditions include  
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
 CC vascular growth. Polypeptides involved with tissue regeneration and  
 CC repair (or nucleic acids encoding them) may be used to promote wound  
 CC healing (e.g., of burns, incisions and ulcers), while those with  
 CC immunomodulatory activities may be used in the treatment of viral,  
 CC bacterial and fungal infections in addition to immune disorders.  
 CC Polypeptides with growth factor activity may be used in cell cultures to  
 CC promote cell growth. For example, such polypeptides may be used to  
 CC manipulate stem cells in culture to give rise to neuroepithelial cells  
 CC that can be used to augment or replace cells damaged by illness,  
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
 CC may also be used in the diagnosis of the above conditions, and in drug  
 CC screening techniques. The present sequence represents a novel human  
 CC polypeptide of the invention.  
 XX  
 XX Sequence 851 AA;

Query Match 100.08; Score 1319; DB 22; Length 851;  
 Best Local Similarity 100.08; Pred. No. 3.8e-98;  
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVGTTDADGEWPMQVSLHALGQGHICGASLISPNLWLSAAHCVIDDRGRYSDDPTOWTA 60  
 DB 611 VVGGTADGEWPMQVSLHALGQGHICGASLISPNLWLSAAHCVIDDRGRYSDDPTOWTA 670  
 QY 61 FLGLHDQSORSGAPGVQERRLRKRIISHPFENFTFDYDIALLELEKPAEYSSMVRPICLPD 120  
 DB 671 flglhdqsrsgapgvqerrlrkriishpfndftfdydiallelelekpaeysmvrpiclpd 730  
 QY 121 ASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVNTTCENLLPQQITPRMVCVGF 180  
 DB 731 ashvfpagkaiwvtgwghtqyggtgallilqkgeirvnttcenllpqgitprmmcvgfl 790  
 QY 181 SGVDSCQDGGPLSSVEADGRIFQAGVSWGDCGCAQRNKGYYTRLPLFRDWIKENTG 240  
 DB 791 sgvdscqdgsgplssveadgrifqagvswgdcgcaqrnkgpyvtrplfrdwikentg 850  
 QY 241 V 241  
 DB 851 V 851  
 RESULT 5  
 AAM25628  
 ID AAM25628 standard; Protein; 851 AA.  
 XX  
 AC AAM25628;  
 XX  
 DT 16-OCT-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:1143.  
 KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
 KW antibacterial; endocrine; cardiant; central nervous system; virucide;  
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
 KW antiagregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;  
 KW dermatological; allergic; antiasthatic; antidiabetic; cytostatic;  
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
 KW neurological disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153455-A2.  
 XX  
 PD 26-JUL-2001.  
 XX  
 PF 22-DEC-2000; 2000WO-US35017.  
 XX  
 PR 23-DEC-1999; 99US-0471275.  
 PR 21-JAN-2000; 2000US-0486725.  
 PR 25-APR-2000; 2000US-0552317.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT;  
 XX  
 XX WPI; 2001-457603/49.  
 DR N-PSDB; AAH99569.  
 XX  
 PT Isolated human polynucleotides encoding polypeptides, useful for the  
 PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
 XX  
 PS Claim 20; Page 237; 1217pp; English.  
 XX  
 CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to  
 CC AAM25963. The proteins can have activities based on the tissues and



Db 855 v 855

RESULT 7

AAB19552

ID AAB19552 standard; Protein; 855 AA.

XX

AC AAB19552;

DT 22-JAN-2001 (first entry)

XX

DE Human matriptase.

XX

KW Matriptase; serine protease; human; breast cancer; pre-malignancy; actinic keratosis; leukoplakia; Barrett's epithelium; columnar metaplasia; ulcerative colitis; Bowenoid papulosis; adenomatous colorectal polyp; Ovary erythroplasia; vulvar intraepithelial neoplasia; tumour; metastasis; therapy.

XX

OS Homo sapiens.

XX

FH Key

FT Region

FT /note= "complement subcomponent 1r and 1s (Clr/s)

FT

FT Region

FT /note= "complement subcomponent 1r and 1s (Clr/s)

FT

FT Region

FT /note= "complement subcomponent 1r and 1s (Clr/s)

FT

FT Region

FT /note= "LDL-receptor type region"

FT

FT Region

FT /note= "LDL-receptor type region"

FT

FT Region

FT /note= "LDL-receptor type region"

FT

FT Region

FT /note= "LDL-receptor type region"

FT

FT Domain

FT /label= Catalytic\_domain

FT

FT Modified-site

FT /note= "N-glycosylated"

FT

FT Modified-site

FT /note= "N-glycosylated"

FT

FT Peptide

FT /note= "RGD motif"

FT

FT Cleavage-site

FT /note= "conserved proteolytic activation site"

XX

PN WO200053232-A1.

XX

XX

XX 14-SEP-2000.

XX

XX 10-MAR-2000; 2000WO-US06111.

XX

XX 12-MAR-1999; 99US-0124006.

XX

XX (GEOU ) UNIV GEORGETOWN.

XX

XX Dickson RB, Lin C, Johnson M, Wang S, Enyedy I;

XX

XX WPI: 2000-594268/56.

XX

XX N-PSDB; AAB88493.

XX

XX Treating malignancies, premalignant and pathologic conditions in a subject, comprises administering matriptase modulating agent -

XX

XX Claim 14; Fig 16; 116pp; English.

XX

XX The present sequence is that of the full-length form of human matriptase, a trypsin-like protease. This has an additional 172 amino acids compared with the truncated form of matriptase given in AAB19551. Either form can be produced in transfected or transfected cells using a claimed method. The zymogen (inactive)

CC form of matriptase is a single-chain protein. The active 2-chain form strongly interacts with fragments of a Kunitz-type serine protease inhibitor (hepatocyte growth factor activator inhibitor, HAI-1) to form SDS-stable complexes. In breast cancer cells, matriptase is present mainly as the uncomplexed form. Only the complexed matriptase is detected in human milk. The invention is directed to a method of detecting a malignancy or a pre-malignant lesion in breast or other tissue by detecting the presence of single- or 2-chain forms of matriptase in the tissue. The object is to inhibit tumour onset, tumour growth and metastasis. CC Malignancies and pre-malignant conditions characterised by expression of the zymogen or activated form of matriptase are treated by administering an inhibitor of matriptase, especially a Bowman-Birk inhibitor. The pre-malignant condition is atypical ductal hyperplasia of the breast, actinic keratosis, leukoplakia, Barrett's epithelium of the esophagus, ulcerative colitis, adenomatous colorectal polyps, erythroplasia of the Queyrat, Bowen's disease, Bowenoid papulosis, vulvar intraepithelial neoplasia or dysplastic changes to the cervix. The invention also provides methods for in vivo or in vitro diagnosis of malignancy or pre-malignant lesion, and methods of identifying matriptase modulators, including activators and inhibitors.

XX

XX Sequence 855 AA;

Query Match 100.0%; Score 1319; DB 21; Length 855;

Best Local Similarity 100.0%; Pred. No. 3.8e-98; Indels 0; Gaps 0;

Matches 241; Conservative 0; Mismatches 0;

QY 1 VVGTDADGEPMQVSLHALGQHICGASLISPNWLVSAAHCYIDRGRYSPTQMTA 60

DB 615 VVGTDADGEPMQVSLHALGQHICGASLISPNWLVSAAHCYIDRGRYSPTQMTA 674

QY 61 FLGLDQSORAPGVQERRLKRISHPPFNDFDYDIALLELEKPAEYSSMVRPICLPD 120

DB 675 FLGLDQSORAPGVQERRLKRISHPPFNDFDYDIALLELEKPAEYSSMVRPICLPD 734

QY 121 ASHVFPAGKAIWVTGWHTQYGGTGALILQKEIRVINQTTCENLFPQITPRMVCYGL 180

DB 735 ASHVFPAGKAIWVTGWHTQYGGTGALILQKEIRVINQTTCENLFPQITPRMVCYGL 794

QY 181 SGVSCQDSCGGLSSVEADGRIFQAGVSWGDCGCAQRNPGVYTRPLPRDWIKENTG 240

DB 795 SGVSCQDSCGGLSSVEADGRIFQAGVSWGDCGCAQRNPGVYTRPLPRDWIKENTG 854

QY 241 V 241

DB 855 V 855

RESULT 8

AAB06930

ID AAB06930 standard; Protein; 855 AA.

XX

XX AAB06930;

XX

DT 16-OCT-2001 (first entry)

XX

XX Human membrane-type serine protease (MTSP) 1.

XX

XX Human; transmembrane serine protease; membrane-type serine protease; MTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic; lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy; matriptase.

XX

XX Homo sapiens.

XX

XX Key

XX Location/Qualifiers

XX Key

XX Domain

XX /label= Protease\_domain

XX

XX 615..855

XX

XX WO200157194-A2.

XX 09-AUG-2001.  
PD  
XX  
XX  
XX 02-FEB-2001; 2001WO-US03471.  
PF  
XX  
XX 03-FEB-2000; 2000US-0179982.  
PR  
XX 18-FEB-2000; 2000US-0183542.  
PR  
XX 22-JUN-2000; 2000US-0213124.  
PR  
XX 26-JUL-2000; 2000US-0220970.  
PR  
XX 08-SEP-2000; 2000US-0657986.  
PR  
XX 22-SEP-2000; 2000US-0234840.  
XX  
XX (CORV-) CORVAS INT INC.  
XX  
XX Madison EL, Ong EO, Yeh J;  
PI  
XX  
XX WPI; 2001-488877/53.  
DR  
XX N-PSDB; AAD13113.  
DR  
XX  
XX Novel single chain polypeptide comprising protease domain of type-II  
PT membrane-type serine protease or its catalytically active portion  
PT useful for treating and preventing cancer and tumor -  
XX  
XX  
XX Claim 12; Page 195-197; 256pp; English.  
PS  
XX  
XX The invention relates to transmembrane serine proteases and their  
CC corresponding nucleotides and the protease domain of a type-II  
CC membrane-type serine protease (MTSP). MTSP is useful for identifying  
CC compounds that modulate or inhibits its proteolytic activity and for  
CC formulating a medicament for treating neoplastic disease. MTSP and  
CC its corresponding nucleotides are useful in preventing or treating  
CC tumors or cancers such as lung carcinoma, colon adenocarcinoma and  
CC ovarian carcinoma, in diagnostics and in hybridisation assays. MTSP  
CC is useful as a diagnostic marker for tumor development, growth and/or  
CC progression and as immunogens to generate antibodies that specifically  
CC bind to it. MTSP DNA is useful in a yeast two-hybrid system and in  
CC gene therapy. The present sequence is human MTSP1 protein (also called  
CC matriptase).  
XX  
XX  
XX Sequence 855 AA;  
SQ

Query Match 100.0%; Score 1319; DB 22; Length 855;  
Best Local Similarity 100.0%; Pred. No. 3.8e-98;  
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VVGGTDADEGEWPQVSLHALGQGHICGASLISPNWLYSAACHYIDDRGFYSPTQWTA 60  
Db 615 vvggtadagewpqvsllhalgqghicgasllspnlwlsaaahcyidrrfysptqwt 674  
QY 61 FLGLHDSQSORAPGVQERLKRRIISHPFNDFTFYDIALLELEKPAEYSSMVRPCLPD 120  
Db 675 flglhdqsrspagvqerrlkrliishpfndftfydiallelekpaeysmvrpiclpd 734  
QY 121 ASHVFAPGKAIWVTGWGHTQYGGTGALLIOLKEIRVINQTCENLLPQGITPRMCMVGF 180  
Db 735 ashvfpagkaiwvtgwghtqygggtgallilqkeirvinqtcenllpqgitprmmcvgfl 794  
QY 181 SGGVDSQCGDSGGLSSVEADGRIFQAGVVSWMGDCGACQARNKPGVYTRLPFRDWMKENTG 240  
Db 795 sggvdsccqsgsgglssveadgrifqagvvs\*wgdcgacqarnkpgvytrlplfrdwikentg 854  
QY 241 V 241  
Db 855 v 855  
RESULT 9  
AAB98500  
ID AAB98500 standard; Protein; 855 AA.  
XX  
XX AAB98500;  
AC  
XX

DT 03-AUG-2001 (first entry)  
XX  
XX Human TAGD-15.  
DE  
XX  
XX Human; TAGD-15; cytostatic; vaccine; ovarian tumour; cancer;  
KW tumour antigen-derived gene 15; extracellular serine protease.  
KW  
XX Homo sapiens.  
CS  
XX  
XX WO200129056-A1.  
PN  
XX  
XX 26-APR-2001.  
PD  
XX  
XX 20-OCT-2000; 2000WO-US29095.  
XX  
XX 20-OCT-1999; 99US-0421213.  
PR  
XX  
XX (UYAR-) UNIV ARKANSAS.  
PA  
XX  
XX O'Brien TJ, Tanimoto H;  
PI  
XX  
XX WPI; 2001-381031/40.  
DR  
XX N-PSDB; AAH23601.  
DR  
XX  
XX Novel extracellular serine protease, termed tumor antigen-derived gene  
PT 15 protein overexpressed in carcinomas and DNA encoding it, for  
PT diagnosis, treatment, prevention of cancer, particularly breast,  
PT ovarian cancer -  
XX  
XX Claim 11; Fig 2; 130pp; English.  
PS  
XX  
XX The present sequence represents human tumour antigen-derived gene 15  
CC (TAGD-15) protein. TAGD-15 is an extracellular serine protease. It was  
CC found that TAGD-15 is over-expressed in ovarian tumours. TAGD-15 protein  
CC or its fragments of 9-20 residues that lack TAGD-15 protease activity are  
CC useful for vaccinating an individual against TAGD-15, having, suspected  
CC of having or at risk of getting cancer. Furthermore, the TAGD-15 gene can  
CC be used as a diagnostic or therapeutic target in cancer.  
XX  
XX  
XX Sequence 855 AA;  
SQ

Query Match 100.0%; Score 1319; DB 22; Length 855;  
Best Local Similarity 100.0%; Pred. No. 3.8e-98;  
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VVGGTDADEGEWPQVSLHALGQGHICGASLISPNWLYSAACHYIDDRGFYSPTQWTA 60  
Db 615 vvggtadagewpqvsllhalgqghicgasllspnlwlsaaahcyidrrfysptqwt 674  
QY 61 FLGLHDSQSORAPGVQERLKRRIISHPFNDFTFYDIALLELEKPAEYSSMVRPCLPD 120  
Db 675 flglhdqsrspagvqerrlkrliishpfndftfydiallelekpaeysmvrpiclpd 734  
QY 121 ASHVFAPGKAIWVTGWGHTQYGGTGALLIOLKEIRVINQTCENLLPQGITPRMCMVGF 180  
Db 735 ashvfpagkaiwvtgwghtqygggtgallilqkeirvinqtcenllpqgitprmmcvgfl 794  
QY 181 SGGVDSQCGDSGGLSSVEADGRIFQAGVVSWMGDCGACQARNKPGVYTRLPFRDWMKENTG 240  
Db 795 sggvdsccqsgsgglssveadgrifqagvvs\*wgdcgacqarnkpgvytrlplfrdwikentg 854  
QY 241 V 241  
Db 855 v 855  
RESULT 10  
AAB35465  
ID AAB35465 standard; Protein; 855 AA.  
XX  
XX AAB35465;  
AC  
XX



DT 06-JUN-2001 (first entry)  
XX Human membrane-type serine protease MT-SPL.  
DE Human; membrane-type serine protease; MT-SPL; cancer.  
XX Homo sapiens.  
XX WO200123524-A2.  
XX PD 05-APR-2001.  
XX 02-OCT-2000; 2000WO-US27250.  
XX 30-SEP-1999; 99US-0410362.  
XX (REGC ) UNIV CALIFORNIA.  
XX Craik CS, Takeuchi T, Shuman M;  
XX WPI; 2001-245002/25.  
XX N-PSDB; AAF28099.  
XX New nucleic acid encoding a membrane type serine protease, useful for  
PT the diagnosis, prognosis and treatment of cancer, particularly  
PT metastatic cancers -  
XX Claim 7; Fig 1; 102pp; English.  
XX The present invention provides the protein and coding sequences for the  
CC novel human membrane-type serine protease MT-SPL. Increased expression of  
CC this protein is associated with cancer, and so the sequences can be used  
CC in cancer diagnosis and the identification of treatments. The present  
CC sequence is the MT-SPL protein.  
XX Sequence 855 AA;  
SQ  
Query Match 100.0%; Score 1319; DB 22; Length 855;  
Best Local Similarity 100.0%; Pred. No. 3.8e-98;  
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VVGTTDADGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFYSDPTQWTA 60  
DB 615 vvggtddadegewpqsylhalgqghicgaslispnwlvsaahcyidrrgrysdptqwtgta 674  
QY 61 FLGLHDSORSAPGVQERRLKRRIISHPFFNDFTFDYDIALLEKPAEYSSMVRPCLPD 120  
DB 675 flglhdqsrspagvqerrlkrliishpfnfdftdydiallelekpaeysmvrpiclpd 734  
QY 121 ASHVFPAGKAIWVTGHTQYGGTGALLIQKEIRVINTTCENLLPQQITPRMCMVGFL 180  
DB 735 ashvfpagkaiwvtgwtgqygtgalllqkgeirvinttcenllpqgitprmmcvgfl 794  
QY 181 SGGVDSGCGDSGGLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYTRPLFRDWIKENTG 240  
DB 795 sggvdsccgdsggplsveadgrifqagvswgdcgcaqrnkpgvytrplfrdwikentg 854  
QY 241 V 241  
DB 855 v 855  
RESULT 11  
AAW22987  
ID AAW22987 standard; Protein; 241 AA.  
XX  
AC AAW22987;  
XX  
DT 08-OCT-1997 (first entry)  
XX Human serine protease 67 (SP67).  
DE  
XX

KW Human; colon carcinoma; COLO 201; cell line; serine protease; SP67;  
XX screening; inhibitor; treatment; disease.  
XX Homo sapiens.  
XX JP09149790-A.  
XX 10-JUN-1997.  
XX 24-JUL-1996; 96JP-0212196.  
XX 29-SEP-1995; 95JP-0275105.  
XX (SUNR ) SUNTORY LTD.  
XX WPI; 1997-357902/33.  
XX N-PSDB; AAT79128.  
XX Human colon carcinoma derived serine protease(s) SP59, SP60 and SP67  
PT - useful to screen for specific inhibitors, e.g. to search for, or  
PT study agent for treatment of various diseases  
XX Claim 1; Pages 12-13; 16pp; Japanese.  
XX The present sequence is the human colon carcinoma COLO 201  
CC cell line derived serine protease 67 (SP67), which can be used to  
CC screen for specific inhibitors, e.g. to search for, or study an  
CC agent for the treatment of various diseases.  
XX Sequence 241 AA;  
SQ  
Query Match 99.7%; Score 1315; DB 18; Length 241;  
Best Local Similarity 99.6%; Pred. No. 2.3e-98;  
Matches 240; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 VVGTTDADGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFYSDPTQWTA 60  
DB 1 vvggtddadegewpqsylhalgqghicgaslispnwlvsaahcyidrrgrysdptqwtv 60  
QY 61 FLGLHDQSRSPAGPVQERRLKRRIISHPFFNDFTFDYDIALLEKPAEYSSMVRPCLPD 120  
DB 61 flglhdqsrspagvqerrlkrliishpfnfdftdydiallelekpaeysmvrpiclpd 120  
QY 121 ASHVFPAGKAIWVTGHTQYGGTGALLIQKEIRVINTTCENLLPQQITPRMCMVGFL 180  
DB 121 ashvfpagkaiwvtgwtgqygtgalllqkgeirvinttcenllpqgitprmmcvgfl 180  
QY 181 SGGVDSGCGDSGGLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYTRPLFRDWIKENTG 240  
DB 181 sggvdsccgdsggplsveadgrifqagvswgdcgcaqrnkpgvytrplfrdwikentg 240  
QY 241 V 241  
DB 241 v 241  
RESULT 12  
ABG21442  
ID ABG21442 standard; Protein; 932 AA.  
XX  
AC ABG21442;  
XX  
DT 18-FEB-2002 (first entry)  
XX Novel human diagnostic protein #21433.  
DE  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
XX WO200175067-A2.  
XX

XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US08631.  
XX 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
XX N-PSDB; AA885629.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity -  
XX Claim 20; SEQ ID No 51801; 103pp; English.  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. AEG00010-ABG30377 represent novel human  
XX diagnostic amino acid sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX Sequence 932 AA;

Query Match 93.0%; Score 1227; DB 22; Length 932;  
Best Local Similarity 93.8%; Pred. No. 1.1e-90;  
Matches 228; Conservative 4; Mismatches 9; Indels 2; Gaps 2;  
Qy 1 VVGTDADGEWQVSLHALGOGHICGASLISPNWLVSAAHCYIDDRGFYSDPTQWT- 59  
Db 590 vvgtdadegewpqlslhalgghicgalslspnwlvsaaahcfidrrfmsyngptqwr 749  
Qy 60 AFLGLHDSQSAP-GVQERRLKRISHPFNFDFDYDIALLELEKPAEYSSMVRPCL 118  
Db 750 aflglhdqsgrrppgwrkrilkrilshpfndftdydiallelelekaeyssmvrpcl 809  
Qy 119 PDASHVPAGKAIWTGNGHTQYGGTGALILQKEIRVINTTCENLLPQQITPMMCVG 178  
Db 810 pdashvpagkaiwtgnghtqygg99tgalllqkeirvlnqtcenllpqqitprmmcvg 869  
Qy 179 FLSGGVDSQSGDGLSSVEADGRIFQAGVVSWMGDGCAQRNKPQVYTRLPFRDWIKEN 238  
Db 870 flsggvdsqsgdglssveadgrifqagvvsweagcvqrnkpgvytrlpfrdwiken 929  
Qy 239 TGV 241  
Db 930 tgv 932  
RESULT 13

AA898507  
ID AAB98507 standard; Protein; 902 AA.  
XX  
AC AAB98507;  
XX  
DT 03-AUG-2001 (first entry)  
XX  
DE Murine epithin.  
XX  
KW Murine; TAGD-15; cytostatic; vaccine; ovarian tumour; cancer; epithin;  
KW tumour antigen-derived gene 15; serine protease.  
XX  
OS Mus musculus.  
XX  
PN WO200129056-A1.  
XX  
PD 26-APR-2001.  
XX  
PF 20-OCT-2000; 2000WO-US29095.  
XX  
PR 20-OCT-1999; 99US-0421213.  
XX  
PA (UYAR-) UNIV ARKANSAS.  
XX  
PI O'Brien TJ, Tanimoto H;  
XX  
DR WPI; 2001-381031/40.  
XX  
XX Novel extracellular serine protease, termed tumor antigen-derived gene  
XX 15 protein overexpressed in carcinomas and DNA encoding it, for  
XX diagnosis, treatment, prevention of cancer, particularly breast,  
XX ovarian cancer -  
XX Disclosure; Page 97-99; 130pp; English.  
XX The present invention relates to human tumor antigen-derived gene 15  
XX (TAGD-15) protein and coding sequence (see AAH23601 and AAB98500).  
XX TAGD-15 is an extracellular serine protease. It was found that TAGD-15 is  
XX over-expressed in ovarian tumours. TAGD-15 protein or its fragments of  
XX 9-20 residues that lack TAGD-15 protease activity are useful for  
XX vaccinating an individual against TAGD-15, having, suspected of having or  
XX at risk of getting cancer. Furthermore, the TAGD-15 gene can be used as a  
XX diagnostic or therapeutic target in cancer. The present sequence was used  
XX in a sequence homology alignment with TAGD-15.  
XX  
SQ Sequence 902 AA;  
Query Match 83.4%; Score 1100; DB 22; Length 902;  
Best Local Similarity 86.9%; Pred. No. 1.8e-80;  
Matches 199; Conservative 17; Mismatches 13; Indels 0; Gaps 0;  
Qy 1 VVGTDADGEWQVSLHALGOGHICGASLISPNWLVSAAHCYIDDRGFYSDPTQWTA 60  
Db 615 vvgtdadegewpqlslhalgghicgalslspnwlvsaaahcfidrrfmsyngptqwr 674  
Qy 61 FLGLHDSQSAP-GVQERRLKRISHPFNFDFDYDIALLELEKPAEYSSMVRPCLPD 120  
Db 675 flglhdqskrsagvqelklrliithpsfndftdydialleleksveystvvrpiclpd 734  
Qy 121 ASHVFPAGKAIWTGNGHTQYGGTGALILQKEIRVINTTCENLLPQQITPMMCVGFL 180  
Db 735 athvfpagkaiwtgnghtqygg99tgalllqkeirvinqtccedlmpqitprmmcvgfl 794  
Qy 181 SGGVDSQSGDGLSSVEADGRIFQAGVVSWMGDGCAQRNKPQVYTRLP 229  
Db 795 sggvdsqsgdglssveadgrifqagvvsweagcvqrnkpgvytrlp 843  
RESULT 14  
ABG21439  
ID ABG21439 standard; Protein; 129 AA.  
XX

AC ABG21439;  
XX 18-FEB-2002 (first entry)  
XX Novel human diagnostic protein #21430.  
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.  
KW Homo sapiens.  
XX WO200175067-A2.  
PN 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US08631.  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
PI WPI: 2001-639362/73.  
XX N-PSDB; AAS85626.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX Claim 20; SEQ ID NO 51798; 103pp; English.  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (II) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX Sequence 129 AA;  
SQ  
Query Match 48.7%; Score 643; DB 22; Length 123;  
Best Local Similarity 92.2%; Pred. No. 1.6e-44;  
Matches 119; Conservative 1; Mismatches 9; Indels 0; Gaps 0;  
QY 106 PNEYSSWRPCLPASHVFPAGKAIWVTGHWYQGGTGLIIQKEIRVINTTCENL 165  
DB 1 pneysswrpclpashvfpagkaiwvtgwhqyggtagllqkgeirvinttce 60  
QY 166 LPQQTTPRMVCGFFSGGVDSQSGPLSVADGRIFQAGVYVWGSDGCAQRNKPgy 225  
DB 61 lpqqltpmmvrgffsggyvdsdsssssspsvveadgrifagvvyvsgdgcqqrnkpgy 120  
QY 226 TRLPFRDW 234  
|||||:|

Db 121 trlpfqdw 129  
RESULT 15  
AAE01942  
ID AAE01942 standard; Protein; 233 AA.  
XX AAE01942;  
XX 31-JUL-2001 (first entry)  
XX Human endotheliase 1 protease domain.  
XX Human; endotheliase 1; protease domain; cytostatic; vulnerary; wound;  
KW neotropic; periodontitis; dermatological disorder; gene therapy; scar;  
KW angiogenesis; cardiovascular disorder; psoriasis; neovascular disease;  
KW chronic inflammatory disease; ocular disorder; circulatory disorder;  
KW crest syndrome; atherosclerosis; haemangiomas; diabetes mellitus;  
KW liver cirrhosis, osteoradionecrosis, systemic sclerosis; oesophageal;  
KW inflammatory bowel disease; fracture; rheumatoid arthritis; retinopathy;  
KW systemic vasculitis; scleroderma; neoplasm; ulcer; burn.  
XX Homo sapiens.  
XX WO200136604-A2.  
XX 25-MAY-2001.  
PD 17-NOV-2000; 2000WO-US31803.  
XX 18-NOV-1999; 99US-0166391.  
PR 22-SEP-2000; 2000US-0234840.  
XX (CORV-) CORVAS INT INC.  
XX Madison EL, Ong EO;  
PI WPI: 2001-336001/35.  
XX N-PSDB; AAD05795.  
XX New nucleic acid encoding a protein comprising endotheliase activity  
PT useful in the prevention and treatment of e.g. vascular malformations,  
PT cardiovascular disorders, and chronic inflammatory disease -  
XX Claim 4; Page 134-135; 152pp; English.  
XX The present sequence is human endotheliase 1 protease domain.  
CC The invention relates to an endotheliase protein, endotheliase protease  
CC domain and their corresponding nucleic acid molecules. An endotheliase  
CC protein or protease domain of it is useful for the treatment and  
CC diagnosis of disorders associated with aberrant angiogenesis or undesired  
CC neovascularisation. The undesired angiogenesis is associated with  
CC disorders selected from solid neoplasm, vascular malformations and  
CC cardiovascular disorders such as angiofibroma, angiolipoma,  
CC atherosclerosis, restenosis/reperfusion injury, arteriovenous  
CC malformations, haemangiomas and vascular adhesions, dyschondroplasia  
CC with vascular hamartomas (Fafucci's syndrome), hereditary haemorrhagic  
CC telangiectasia (Rendu-Osler-Weber syndrome) and Von Hippel Lindau  
CC syndrome, chronic inflammatory diseases such as diabetes mellitus,  
CC haemophilic joints, inflammatory bowel disease, nonhealing fractures,  
CC periodontitis, psoriasis, rheumatoid arthritis, venous stasis ulcers,  
CC granulations/burns, hypertrophic scars, liver cirrhosis,  
CC osteoradionecrosis, postoperative adhesion, pyogenic granuloma and  
CC systemic sclerosis and aberrant wound repairs, circulatory disorders  
CC Raynaud's phenomenon, crest syndromes such as calcinosis, oesophageal,  
CC dysmotility, sclerodactyly and telegiectasis, dermatological disorders  
CC such as systemic vasculitis, scleroderma, pyoderma gangrenosum,  
CC vasculopathy, venous, arterial ulcers, Sturge-Weber syndrome, Port-wine  
CC stains, blue rubber bleb nevus syndrome, Klippel-Trenaunay-Weber syndrome  
CC and Osler-Weber-Rendu syndrome and ocular disorders such as blindness  
CC caused by ocular neovascular disease, corneal graft neovascularisation,  
CC macular degeneration, retinopathy of prematurity, retrolental  
CC fibroplasia and corneal neovascularisation. The nucleic acids of the



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2002, 17:08:24 ; Search time 19.03 Seconds  
(without alignments)  
1216.896 Million cell updates/sec

Title: US-09-657-986b-2

Perfect score: 1319

Sequence: 1 VVGTDADGEWPQVSLHA.....PGVYTRLPFRDWIKENTGV 241

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_71:\*

1: Pir1:\*

2: Pir2:\*

3: Pir3:\*

4: Pir4:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1151	87.3	855	2 JC7731	membrane-bound arg
2	1151	87.3	855	2 JC7775	membrane type-seri
3	552.5	41.9	1034	1 A53663	enteropeptidase (E
4	547.5	41.5	1019	1 A56318	enteropeptidase (E
5	544.5	41.3	1035	1 A43090	enteropeptidase (E
6	483	36.6	786	1 A47547	serine proteinase
7	477	36.2	638	1 KOMSPL	plasma kallikrein
8	473.5	35.9	417	1 S0845	hepsin (EC 3.4.21.
9	472	35.8	790	1 PLPG	plasmin (EC 3.4.21
10	469	35.6	638	1 KORTPL	plasma kallikrein
11	467	35.4	275	2 S40007	trypsin (EC 3.4.21
12	466	35.3	416	1 S33777	hepsin (EC 3.4.21.
13	463.5	35.1	1524	2 T30337	polyprotein - Afri
14	462	35.0	275	2 S40005	trypsin (EC 3.4.21
15	462	35.0	638	1 K0HUP	plasma kallikrein
16	461	35.0	343	1 A57014	proctasin (EC 3.4.
17	457	34.6	274	2 S35339	trypsin (EC 3.4.21
18	456.5	34.6	267	2 S40006	trypsin (EC 3.4.21
19	455	34.5	625	1 KFHU1	coagulation factor
20	453.5	34.4	812	1 PLMS	plasmin (EC 3.4.21
21	452	34.3	460	2 B61545	plasmin (EC 3.4.21
22	452	34.3	810	1 PLHU	plasmin (EC 3.4.21
23	448.5	34.0	810	2 146260	plasmin (EC 3.4.21
24	446	33.8	238	1 TRWV5Y	trypsin-like prote
25	444	33.7	455	2 A61545	plasmin (EC 3.4.21
26	441	33.4	810	2 B30848	plasmin (EC 3.4.21
27	439.5	33.3	812	1 PLBO	plasmin (EC 3.4.21
28	439	33.3	270	2 S56160	mast cell tryptase
29	436.5	33.1	416	1 KFB0	coagulation factor

#### ALIGNMENTS

RESULT 1

JC7731

membrane-bound arginine-specific serine proteinase precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 14-Dec-2001

C:Accession: JC7731

R:Kishi, K.; Yamazaki, K.; Yasuda, I.; Yahagi, N.; Ichinose, M.; Tsuchiya, Y.; Athaud

J. Biochem. 130, 425-430, 2001

A:Title: Characterization of a membrane-bound arginine-specific serine protease from

A:Reference number: JC7731; MUID:21421307; PMID:11530019

A:Accession: JC7731

A:Molecule type: mRNA

A:Residues: 1-855 <KIS>

A:Cross-references: DDBJ:AB049189

A:Experimental source: strain Male, 7-week-old

C:Comment: This enzyme, localized mainly on brushborder membranes of the intestine, p

C:Keywords: protein digestion

Query Match 87.3%; Score 1151; DB 2: Length 855;

Best Local Similarity 87.1%; Pred. No. 7,4e-99;

Matches 210; Conservative 15; Mismatches 16; Indels 0; Gaps 0;

QY 1 VVGTDADGEWPQVSLHALCGGHICGASLISPNWLSAAHCYIDDRGRYSPTQWTA 60

Db 615 VVGTDADGEWPQVSLHALCGGHICGASLISPNWLSAAHCYIDDRGRYSPTQWTA 674

QY 61 FLGLHDQSORSPAGVQERLKRILSHPPFNDFDYDIALLEKPAEYSSMVRICLPD 120

Db 675 FLGLHDQSORSPAGVQERLKRILSHPPFNDFDYDIALLEKPAEYSSMVRICLPD 734

QY 121 ASHVFPAGKAIWVTGNGHTQYGGTGALILQKGEIRVINOTTCEILLPOQITPRMVCVGL 180

Db 735 NTHVFPAGKAIWVTGNGHTQYGGTGALILQKGEIRVINOTTCEILLPOQITPRMVCVGL 794

QY 181 SGVDSCDGSGLSSVEADGRIFQAGVSWGDCGACORNPQVYTRLPFRDWIKENTG 240

Db 795 SGVDSCDGSGLSSVEADGRIFQAGVSWGDCGACORNPQVYTRLPFRDWIKENTG 854

QY 241 V 241

Db 855 V 855

RESULT 2

JC7775

membrane type-serine protease 1 - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002

C:Accession: JC7775

R:Satom, S.; Yamazaki, Y.; Tsuzuki, S.; Hitomi, Y.; Iwanaga, T.; Fushiki, T.

Biochem. Biophys. Res. Commun. 287, 995-1002, 2001



F:785-1019/Product: enteropeptidase light chain #status predicted <LCH>  
 F:785-1014/Domain: trypsin homology <TRY>  
 F:116-147/179,328,335,388,440,470,503,534,630,682,706,725,848,887,909,949/Binding site:  
 F:772-896,810-826,910-977,941-956,967-995/Disulfide bonds: #status predicted  
 F:825,876,971/Active site: His, Asp, Ser #status predicted

Query Match 41.5%; Score 547.5; DB 1; Length 1019;  
 Best Local Similarity 44.7%; Pred. No. 1.2e-42;  
 Matches 106; Conservative 44; Mismatches 80; Indels 7; Gaps 4;

QY 1 VVGTDADGEWMPQVSLHALCGGHICGASLISPNWLVSAAHCYIDDRGFYSPTQWTA 60  
 DB 785 IVGGSNAKGCAMPWVGLY-YGRLICGLASLVSSDLVLSAAHCYVYV-----RNLEPSKWT 839  
 QY 61 FLGLHDSQSRAPGVQERRLKRIISHPFNFDFDYDIALLEKPAEYSSWVRPCLPD 120  
 DB 840 ILGLHKNLSPTQVPRIDEIVINPHNRRKNDLMMHLEKFNVTYDIQICLPE 899  
 QY 121 ASHVFPAGKAIWVTGHTQYGGTGALILQKEIRVINOTTENLLPO-QITPRMCMVGF 179  
 DB 900 ENQVFPGRNCSTAGTGVYOGTANILQEADVPLLSNERCQQQMPYNTENNICAGY 959  
 QY 180 LSGGVSCDGGSGPLSSVEADGRIFQAGVSWGDCGCAORNKPGVYTRPLFRDWIK 236  
 DB 960 EEGGIDSCDGGSGPLMCOE-NNRWFLAGVTSGYKCALPNRPGVIARVSRFTWQ 1015

RESULT 5  
 A43090  
 N:Enteropeptidase (EC 3.4.21.9) precursor - bovine  
 C:Species: Bos primigenius taurus (cattle)  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: A43090; A48874; A61436  
 R:Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994  
 A:Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease composed of two distinct regions  
 A:Reference number: A43090; MUID:94329561  
 A:Accession: A43090  
 A:Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1035 <KIT>  
 A:Cross-references: GB:U09859; NID:g746410; PIDN:AAB40026.1; PID:g746411  
 A:Experimental source: small intestine  
 R:LaValle, E.R.; Rehmtulla, A.; Racie, L.A.; DiBlasio, E.A.; Ferenz, C.; Grant, K.L.;  
 J. Biol. Chem. 268, 23311-23317, 1993  
 A:Title: Cloning and functional expression of a cDNA encoding the catalytic subunit of human enterokinase  
 A:Reference number: A48874; MUID:94043122  
 A:Accession: A48874  
 A:Molecule type: mRNA  
 A:Residues: 801-1035 <LAV>  
 A:Cross-references: GB:L19663; NID:g416131; PIDN:AAA16035.1; PID:g416132  
 A:Note: parts of this sequence, including the amino end of the mature protein, were confirmed by N-terminal sequencing  
 R:Light, A.; Janska, H.  
 J. Protein Chem. 10, 475-480, 1991  
 A:Title: The amino-terminal sequence of the catalytic subunit of bovine enterokinase.  
 A:Reference number: A61436; MUID:92189715  
 A:Accession: A61436  
 A:Molecule type: protein  
 A:Residues: 801-807, 'y', 809-827 <LIG>  
 C:Comment: The mechanism of association with the membrane of the intestinal brush border membrane attachment using a signal-anchor sequence.  
 C:Comment: Conversion from membrane-bound to soluble forms may involve further processing  
 C:Complex: mature enteropeptidase is variously reported to contain two (heavy and light) lflde linked  
 C:Function:  
 A:Description: cleaves propeptide from trypsinogen to produce active trypsin  
 A:Pathway: intestinal digestive hydrolase cascade  
 C:Superfamily: enteropeptidase; Ctr/Cls repeat homology; LDL receptor ligand-binding repeat  
 C:Keywords: glycoprotein; hydrolase; intestine; serine proteinase; transmembrane protein  
 F:22-38/Domain: transmembrane #status predicted <TM>  
 F:52-117/Product: enteropeptidase mini chain #status predicted <MCH>

F:118-800/Product: enteropeptidase heavy chain #status predicted <HCH>  
 F:199-236/Domain: LDL receptor ligand-binding repeat homology <LDL>  
 F:358-520/Domain: MAM homology <MAM>  
 F:542-647/Domain: Ctr/Cls repeat homology <CTR>  
 F:559-693/Domain: LDL receptor ligand-binding repeat homology <LDL2>  
 F:594-799/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SCR>  
 F:801-1035/Product: enteropeptidase light chain #status predicted <LCH>  
 F:801-1030/Domain: trypsin homology <TRY>  
 F:116-147,170,194,233,263,264,404,456,486,519,550,646,698,722,741,762,864,903,965/Bln  
 F:788-912,826-842,926-993,957-972,983-1011/Disulfide bonds: #status predicted  
 F:841,892,987/Active site: His, Asp, Ser #status predicted

Query Match 41.3%; Score 544.5; DB 1; Length 1035;  
 Best Local Similarity 43.5%; Pred. No. 2.3e-42;  
 Matches 103; Conservative 50; Mismatches 77; Indels 7; Gaps 4;

QY 1 VVGTDADGEWMPQVSLHALCGGHICGASLISPNWLVSAAHCYIDDRGFYSPTQWTA 60  
 DB 801 IVGGSNREGAMPWVVALYFDQ-QVCGASLVSRDLVLSAAHCYVYV-----RNMEPSKWA 855  
 QY 61 FLGLHDSQSRAPGVQERRLKRIISHPFNFDFDYDIALLEKPAEYSSWVRPCLPD 120  
 DB 856 VLGLHMASNLTSPTQVPRIDEIVINPHNRRKNDLMMHLEKFNVTYDIQICLPE 915  
 QY 121 ASHVFPAGKAIWVTGHTQYGGTGALILQKEIRVINOTTENLLPO-QITPRMCMVGF 179  
 DB 916 ENQVFPGRNCSTAGTGVYOGTANILQEADVPLLSNERCQQQMPYNTENNICAGY 975  
 QY 180 LSGGVSCDGGSGPLSSVEADGRIFQAGVSWGDCGCAORNKPGVYTRPLFRDWIK 236  
 DB 976 EAGGVSCDGGSGPLMCOE-NNRWLLAGVTSGYQCALPNRPGVIARVSRFTWQ 1031

RESULT 6  
 A47547  
 C:Species: Drosophila melanogaster  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: A47547  
 R:Appel, L.F.; Prout, M.; Abu-Shumays, R.; Hammonds, A.; Garbe, J.C.; Fristrom, D.; F  
 Proc. Natl. Acad. Sci. U.S.A. 90, 4937-4941, 1993  
 A:Title: The Drosophila Stubble-stubloid gene encodes an apparent transmembrane serine  
 A:Reference number: A47547; MUID:93281671  
 A:Accession: A47547  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-786 <APP>  
 A:Cross-references: GB:L11451; NID:g158511; PIDN:AAA28918.1; PID:g158512  
 C:Genetics:  
 A:Gene: Sb-sbd  
 A:Cross-references: FlyBase:FBgn000319  
 C:Superfamily: serine proteinase stubble-stubloid; trypsin homology  
 C:Keywords: hydrolase; serine proteinase; transmembrane protein  
 F:61-77/Domain: transmembrane #status predicted <TM>  
 F:543-781/Domain: trypsin homology <TRY>

Query Match 36.6%; Score 483; DB 1; Length 786;  
 Best Local Similarity 39.2%; Pred. No. 8.9e-37;  
 Matches 98; Conservative 47; Mismatches 85; Indels 20; Gaps 8;

QY 1 VVGTDADGEWMPQVSLH-----ALGQGHICGASLISPNWLVSAAHCYIDDRGFYSDP 55  
 DB 543 IVGGSAAFGRPQVSVRSTFFGSFTRCGGALINENNIATAGHC-VDDLII-----596  
 QY 56 TQWTAFLGHDS--ORSAPGVQERRLKRIISHPFNFDFDYDIALLEKPAEYSSWV 113  
 DB 597 SQIRIVGVDISHVQEQUPYERGVAKV--HPKISFETYEYDIALVLEKPLEFAPHV 655  
 QY 114 RPLCPDASHVFPAGKAIWVTGHTQYGGTGALILQKEIRVINOTTENLL-----PQQ 169  
 DB 656 SPICLETDSLL-IGNNATVTCWGRSLSEGGTLPVSLQEVSPVIVSNDNCKSMENRAGQE 714

QY 170 ITPRM-MCVGFLSGVDSCQGDSGPLSSVEADRIQTQAGVWSWGDCGAQRNPGVVTRL 228  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 715 FIPDFLCAGYETGGSDCGDGLPQAQSODGRFFLAGIISWIGCAEANLPGVCTRI 774

QY 229 PLFRDWIKEN 238  
| : | : | :  
Db 775 SKFTPWILEH 784

RESULT 7  
KOMSP  
plasma kallikrein (EC 3.4.21.34) precursor - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 18-Jun-1999  
R:Accession: A36557  
R:Seidah, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaubien, G.; Brachpapa, L.; Rochemond  
DNA Cell Biol. 9, 737-748, 1990  
A>Title: Mouse plasma kallikrein; cDNA structure, enzyme characterization, and compariso  
A:Reference number: A36557; MUID:91090844  
A:Accession: A36557  
A:Molecule type: mRNA  
A:Residues: 1-638 <SEI>  
A:Cross-references: GB:M58588; NID:g200358; PIDN:AAA63393.1; PID:g200359  
A>Note: part of this sequence, including the amino ends of both the heavy and light chain  
C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w  
are linked by one or more disulfide bonds.  
C:Superfamily: coagulation factor XI; trypsin homology  
K:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla  
F:19/Domain: signal sequence #status predicted <SIG>  
F:20-300/Product: plasma kallikrein heavy chain #status experimental <HC>  
F:20-109/Domain: apple repeat <AP1>  
F:110-199/Domain: apple repeat <AP2>  
F:200-289/Domain: apple repeat <AP3>  
F:291-380/Domain: apple repeat <AP4>  
F:391-638/Product: plasma kallikrein light chain #status experimental <LC>  
F:391-621/Domain: trypsin homology <TRY>  
F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322  
F:127,215,308,396,494/Binding site: Carbohydrate (Asn) (covalent) #status predicted  
F:434,483,578/Active site: His, Asp, Ser #status predicted

```

Query Match      36.2%; Score 477; DB 1; Length 638;
Best Local Similarity 39.3%; Pred. No. 2.5e-36;
Matches 96; Conservative 43; Mismatches 91; Indels 14; Gaps 7;

QY   1 VVGGTDADEGENPQVSLHA--LGGHICGASLISPNWLVSAAHCYIDDRGFYSDDPTQW 58
       :|||: ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   391 IVGGTNASLGEPWQVSLQVLVSQTHLCGSGIIGROWLTAAHCF--DGIPY--PDYW 445
       : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY   59 TAFGLGHDSQ--RSAPGVQERRLKRIISHPPFNDFDYDIALLELEKPAEYSSMVRED 116
       : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db   446 RIYGGILSEITKEFP--SSRIKELIHQEVKSEGNDYDIALIKQLTPLNTEFEQRDI 502
       : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY   117 CLPDASHVPFAGKALVTWGHTGYGTTGAILLOKEGIRVINOTTCENLLPOO-IIPRMM 175
       : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db   503 CLPSKADTNTIYTNCWVTGWGTYKEQGETQNILQXATPLVPNEEQCKYRDVINKQMI 562
       : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY   176 CVGFLSGGVDSQQSGSGLSPSVREADGRIFQAGVWSNGDCGARNPKQGVYTRLPJFRDMI 235
       : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db   563 CAGYKEGGTDACKGSDGGPL-VCKHSRGRWOLVIGITSWGEGCGRKDPGVYTKVSYEMDWI 621
       : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY   236 KENT 239
       || |
Db   622 LEXT 625

RESULT      8
S00845
hepsin (EC 3.4.21.-) - human
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1988 #sequence revision 31-Dec-1988 #text change 18-Jun-1999
```

C:Accession: S00845  
R:Leytus, S.P.; Loeb, K.R.; Hagen, F.S.; Kurachi, K.; Davie, E.W.  
Biochemistry 27, 1067-1074, 1988  
A:Title: A novel trypsin-like serine protease (hepsin) with a putative transmembrane  
A:Reference number: S00845; MUID:88209431  
A:Accession: S00845  
A:Molecule type: mRNA  
A:Residues: 1-417 <LEV>  
A:Cross-references: EMBL:X07732; NID:g32063; PIDN:CAA30558.1; PID:g32064  
C:Genetics:  
A:Gene: GDB:HPN; TMRPSL1; hepsin  
A:Cross-references: GDB:135685; OMIM:142440  
A:Map position: 19q11-19q13.2  
C:Superfamily: hepsin; trypsin homology  
C:Keywords: hydrolase; liver; serine proteinase; transmembrane protein  
F:23-45/Domain: transmembrane #status predicted <TMN>  
F:163-400/Domain: trypsin homology <TRY>  
F:188-204, 291-359, 322-338, 349-381/disulfide bonds: #status predicted  
F:203, 257, 353/Active site: His, Asp, Ser #status predicted

Query Match	35.9%	Score	473.5	DB 1	Length	417			
Best Local Similarity	40.7%	Pred. No.	3.2e-36						
Matches	100	Conservative	38	Mismatches	89	Indels	19	Gaps	8
Qy	1	VVGDTADGEQWQVSLHALGGHICGASLISPNWLYSAACHYCIDRGFRYSDPTQWTA	60						
Db	163	IVGGRTSLGRWQVWVSLYDQ-AHLCGSLLSGDWLTAHCFPE---	RNRVLSRWV	217					
Qy	61	FLGLHQSQRSAPGVQERKLRIISH----	PPF--NDTFDYDIALLEKPAEYSSMW	114					
Db	218	FAGA--VAQASPHGLQ-LGVQAVVHGGLPFPRPNSENSNDIALVHLSPLPTEYIQ	274						
Qy	115	PICLPDASHVFPAGKAIWTTGWGHTQYGGTQALILQKEIRVINQTTCE--	NLIPQOITP	172					
Db	275	PVCLPAAGQALVDGKICTVTGNGNTYYGQAGVLQEARVPIISNDVCGNADFEVGNQIKP	334						
Qy	173	RMCMVGLSGVDSCQDGGGL----	SSVEADGRIFQAGVYVSGDGCQAQRNKGVYTRLP	229					
Db	335	KMFCAGTPEGGIDACQDGGGPFVCEDSISRTPRWRLCGIVSWGTCGALCAQKPGVITKVS	394						
Qy	230	LFRDWI	235						
Db	395	DFREWI	400						

```

RESULT      9
PLPG
plasmin (EC 3.4.21.7) precursor - pig (fragment)
N:Alternate names: plasminogen
N:Contains: miniplasminogen
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 07-Sep-1990 #sequence_revision 01-Nov-1996 #text_change 18-Jul-1997
C:Accession: S03733; S03737; A25834
R:Schaller, J.; Marti, T.; Roessellet, S.J.; Kaempfer, U.; Rickli, E.E.
Fibrinolysis 1, 91-102, 1987
A:Title: Amino acid sequence of the heavy chain of porcine plasmin. Comparison of the
A:Reference number: S03733
A:Accession: S03733
A:Molecule type: protein
A:Residues: 1-560 <SCH>
R:Brunisholz, R.A.; Lerch, P.G.; Schaller, J.; Rickli, E.E.; Lergier, W.; Manneberg,
Eur. J. Biochem. 114, 465-470, 1981
A:Title: Comparison of the primary structure of the N-terminal CNB fragments of human
A:Reference number: S03735; MUID:81212097
A:Accession: S03737
A:Molecule type: protein
A:Residues: 1-57 <BRU>
R:Marti, T.; Schaller, J.; Rickli, E.E.
Eur. J. Biochem. 149, 279-285, 1985
A:Title: Determination of the complete amino-acid sequence of porcine miniplasminogen
A:Reference number: A25834; MUID:85203907
A:Accession: A25834

```



[illegible]





---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2002, 17:10:35 ; Search time 11.03 Seconds

(without alignments)  
846.003 Million cell updates/sec

Title: US-09-657-986B-2

Perfect score: 1319

Sequence: 1 VVGSTDADEGEWPMQVSLHA.....PGVYTRPLFRDMIKENTGV 241

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt.40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1319	100.0	855	1 ST14_HUMAN	Q9Y5Y6 homo sapien
2	1152	87.3	855	1 ST14_MOUSE	P56677 mus musculu
3	576.5	43.7	422	1 DES1_HUMAN	Q9U152 homo sapien
4	552.5	41.9	1034	1 ENTK_PIG	P98074 sus scrofa
5	547.5	41.5	1019	1 ENTK_HUMAN	P98073 homo sapien
6	544.5	41.3	1035	1 ENTK_BOVIN	P98072 bos taurus
7	510	38.7	490	1 TMS2_MOUSE	Q9J1Q8 mus musculu
8	503	38.1	1069	1 ENTK_MOUSE	P97435 mus musculu
9	497.5	37.7	418	1 HATT_HUMAN	O60235 homo sapien
10	497	37.7	492	1 TMS2_HUMAN	O15393 homo sapien
11	483	36.6	455	1 TMS5_MOUSE	Q9SR04 mus musculu
12	483	36.6	786	1 STUB_DROME	Q03319 drosophila
13	481.5	36.5	416	1 HEP5_MOUSE	O38453 mus musculu
14	477	36.2	457	1 TMS5_HUMAN	Q9H3S3 homo sapien
15	477	36.2	638	1 KAL_MOUSE	P26262 mus musculu
16	474.5	36.0	454	1 TMS3_HUMAN	P57727 homo sapien
17	473.5	35.9	417	1 HEP5_HUMAN	P05981 homo sapien
18	472	35.8	790	1 PLMN_PIG	P06867 sus scrofa
19	471.5	35.7	290	1 MPN_HUMAN	Q9BGR3 homo sapien
20	469	35.6	638	1 KAL_RAT	P14272 rattus norv
21	467	35.4	275	1 TRY3_ANOGA	P35037 anopheles g
22	466	35.3	416	1 HEP5_RAT	Q05511 rattus norv
23	462	35.0	275	1 TRY4_ANOGA	P35038 anopheles g
24	462	35.0	638	1 KAL_HUMAN	P03952 homo sapien
25	461	35.0	343	1 PSS8_HUMAN	Q16651 homo sapien
26	460	34.9	275	1 TRY7_PIG	Q9A2D1 sus scrofa
27	457	34.6	342	1 PSS8_RAT	Q9S87 rattus norv
28	456.5	34.6	267	1 TRY7_ANOGA	P35041 anopheles g
29	455	34.5	625	1 FAIL_HUMAN	P03951 homo sapien
30	454	34.4	274	1 TRY1_ANOGA	P35035 anopheles g
31	453.5	34.4	812	1 PLMN_MOUSE	P20918 mus musculu
32	452	34.3	343	1 PLMN_SHEEP	P81286 ovis aries
33	452	34.3	810	1 PLMN_HUMAN	P00747 homo sapien

ALIGNMENTS

RESULT	ID	ST14_HUMAN	STANDARD;	PRT;	855 AA.
34	448.5	34.0	810	1	PLMN_ERIEU
35	447	33.9	333	1	PLMN_CANFA
36	446	33.8	238	1	TRY5_AEDAE
37	444	33.7	338	1	PLMN_HORSE
38	442	33.5	342	1	PSS8_MOUSE
39	441	33.4	810	1	PLMN_MACMU
40	441	33.4	1042	1	CORI_HUMAN
41	439.5	33.3	812	1	PLMN_BOVIN
42	439	33.3	270	1	TRY7_MERON
43	437	33.1	273	1	TRY7_SHEEP
44	436.5	33.1	416	1	FA9_BOVIN
45	436	33.1	254	1	TRY3_AEDAE
34	448.5	34.0	810	1	PLMN_ERIEU
35	447	33.9	333	1	PLMN_CANFA
36	446	33.8	238	1	TRY5_AEDAE
37	444	33.7	338	1	PLMN_HORSE
38	442	33.5	342	1	PSS8_MOUSE
39	441	33.4	810	1	PLMN_MACMU
40	441	33.4	1042	1	CORI_HUMAN
41	439.5	33.3	812	1	PLMN_BOVIN
42	439	33.3	270	1	TRY7_MERON
43	437	33.1	273	1	TRY7_SHEEP
44	436.5	33.1	416	1	FA9_BOVIN
45	436	33.1	254	1	TRY3_AEDAE
AC	Q9Y5Y6;	Q9H3S0;	Q9HCA3;	Q9BS01;	Q9HB36;
DT	16-OCT-2001	(Rel. 40,	Created)		
DT	16-OCT-2001	(Rel. 40,	Last sequence update)		
DE	Suppressor of tumorigenicity 14	(EC 3.4.21.-)	(Matriptase)		(Membrane-
DE	type serine protease 1) (MT-Sp1)	(Prostamin)	(Serine protease TADG-15)		(Tumor associated differentially-expressed gene-15 protein).
GN	ST14 OR PRSS14 OR SNCL9 OR TADG15.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
NCBI_TaxID=9606;					
SEQUENCE FROM N.A.					
MEDLINE=99303581;	PubMed=10373424;				
Lin C.Y., Anders J., Johnson M., Sang O.A., Dickson R.B.;					
"Molecular cloning of cDNA for matriptase, a matrix-degrading serine					
protease with trypsin-like activity."					
J. Biol. Chem. 274:18231-18236(1999).					
SEQUENCE FROM N.A.					
MEDLINE=99432178;	PubMed=10500122;				
Takeuchi T., Shuman M.A., Craik C.S.;					
"Reverse biochemistry: Use of macromolecular protease inhibitors to					
dissect complex biological processes and identify a membrane-type					
serine protease in epithelial cancer and normal tissue."					
Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).					
SEQUENCE FROM N.A.					
TISSUE=Prostate;					
Yamaguchi N., Mitsui S.;					
"Molecular cloning of a novel transmembrane serine protease expressed					
in human prostate."					
Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.					
SEQUENCE FROM N.A.					
Tanimoto H., Underwood L.J., Wang Y., Shigemasa K., Parmley T.H.,					
O'Brien T.J.;					
Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.					
SEQUENCE OF 327-855 FROM N.A.					
TISSUE=Muscle;					
Strausberg R.;					
Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.					
SEQUENCE OF 340-664 FROM N.A.					
Cao J., Fan W., Zheng S.;					
"Genomic analysis of a novel human serine protease SNCL9."					
Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.					
CHARACTERIZATION.					
TISSUE=Milk;					
MEDLINE=99303582;	PubMed=10373425;				

Lin C.Y., Anders J., Johnson M., Dickson R.B.;  
 "Purification and characterization of a complex containing matrilysin  
 and a Kunitz-type serine protease inhibitor from human milk.";  
 J. Biol. Chem. 274:18237-18242(1999).  
 CC -!- FUNCTION: DEGRADERS EXTRACELLULAR MATRIX. PROPOSED TO PLAY A ROLE  
 CC IN BREAST CANCER INVASION AND METASTASIS. EXHIBITS TRYPSIN-LIKE  
 CC ACTIVITY AS DEFINED BY CLEAVAGE OF SYNTHETIC SUBSTRATES WITH ARG  
 CC OR LYS AS THE P1 SITE.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).  
 CC -!- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF118224; AAD42765.2; -  
 CC EMBL; AF133086; AAF00109.1; -  
 CC EMBL; AB030036; BAB20376.1; -  
 CC EMBL; AF057145; AAG15395.1; -  
 CC EMBL; BC005826; AAH05826.1; -  
 CC EMBL; AF283256; AAG13949.1; -  
 CC HSP; P00763; IDPO.  
 CC MEROPS; S01.302; -  
 CC InterPro; IPR000859; CUB.  
 CC InterPro; IPR001314; Chymotrypsin.  
 CC InterPro; IPR002172; LDL\_recept\_A.  
 CC InterPro; IPR001254; Trypsin.  
 CC Pfam; PF00431; CUB; 2. Trypsin.  
 CC Pfam; PF00057; ldl\_recept\_a; 4.  
 CC Pfam; PF00089; trypsin; 1.  
 CC PRINTS; PR00722; CHYMOTRYPSIN.  
 CC PRINTS; PR00261; LDLRECEPTOR.  
 CC SMART; SM00042; CUB; 2.  
 CC SMART; SM00192; LDLa; 3.  
 CC SMART; SM00020; Tryp\_SPC; 1.  
 CC PROSITE; PS01180; CUB; 2.  
 CC PROSITE; PS01209; LDLR\_1; 2.  
 CC PROSITE; PS00068; LDLR\_2; 4.  
 CC PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 CC PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 CC PROSITE; PS00135; TRYPSIN\_SER; 1.  
 CC Signal-anchor; Glycoprotein; Hydrolase; Serine protease;  
 CC Transmembrane; Repeat.  
 CC DOMAIN 1 55  
 CC TRANSMEM 56 76  
 CC  
 CC CYTOPLASMIC (POTENTIAL).  
 CC SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 CC (POTENTIAL).  
 CC EXTRACELLULAR (POTENTIAL).  
 CC CUB 1.  
 CC CUB 2.  
 CC LDL-RECEPTOR CLASS A 1.  
 CC LDL-RECEPTOR CLASS A 2.  
 CC LDL-RECEPTOR CLASS A 3.  
 CC LDL-RECEPTOR CLASS A 4.  
 CC SERINE PROTEASE.  
 CC CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CC -CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CC -CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC PEA -> GTR (IN REF. 5).  
 CC R -> S (IN REF. 4).  
 CC A -> V (IN REF. 3).  
 CC  
 CC 855 AA; 94769 MW; 26143132C0IF99C9 CRC64;  
 CC SEQUENCE

Query Match 100.0%; Score 1319; DB 1; Length 855;  
 Best Local Similarity 100.0%; Pred. No. 6.2e-117;  
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VVGSTDADEGEWQVSLHALGQGHICGASLISPNVLVSAHCVIIDDGRFYSDDPTQWTA 60  
 DB 615 VVGSTDADEGEWQVSLHALGQGHICGASLISPNVLVSAHCVIIDDGRFYSDDPTQWTA 674  
 QY 61 FLGLHDQSQRSAPGQVQERRLKRIISHFFNFDTFDYDIALLELEKPAEYSSMVRPICLPD 120  
 DB 675 FLGLHDQSQRSAPGQVQERRLKRIISHFFNFDTFDYDIALLELEKPAEYSSMVRPICLPD 734  
 QY 121 ASHVPPAGKAIWITGWHGTQVGGTGALLQKGEIRVINQTCENLLPQQTIPRMVGVGL 180  
 DB 735 ASHVPPAGKAIWITGWHGTQVGGTGALLQKGEIRVINQTCENLLPQQTIPRMVGVGL 794  
 QY 181 SGGVSCQGDGSGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYTRLPFRWIKENTG 240  
 DB 795 SGGVSCQGDGSGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYTRLPFRWIKENTG 854  
 QY 241 V 241  
 DB 855 V 855  
 RESULT 2  
 ID ST14\_MOUSE STANDARD; PRT: 855 AA.  
 AC P56677;  
 DT 15-JUL-1999 (Rel. 38, Created);  
 DT 16-OCT-2001 (Rel. 40, Last sequence update);  
 DT 16-OCT-2001 (Rel. 40, Last annotation update);  
 DE Suppressor of tumorigenicity 14 (EC 3.4.21.-) (Epithin).  
 GN ST14 OR PRSS14.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C.B.17SCID; TISSUE=Thymus;  
 RX MEDLINE=99216440; PubMed=10199918;  
 RA Kim M.G., Chen C., Lyu M.S., Cho E.G., Park D., Kozak C.,  
 RA Schwartz R.H.;  
 RT "Cloning and chromosomal mapping of a gene isolated from thymic  
 RT stromal cells encoding a new mouse type II membrane serine protease,  
 RT epithin, containing four LDL receptor modules and two CUB domains.";  
 RL Immunogenetics 49:420-428(1999).  
 RN [2]  
 RP REVIEWS TO 23; 321; 325; 343; 409-410 AND C-TERMINUS.  
 RC STRAIN=C.B.17SCID; TISSUE=Thymus;  
 RA Kim M.G., Chen C., Cho E.G., Park D., Schwartz R.H.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Breast tumor;  
 RA Strausberg R.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).  
 CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN INTESTINE, KIDNEY, LUNG,  
 CC AND THYMUS. NOT EXPRESSED IN SKELETAL MUSCLE, LIVER, HEART,  
 CC TESTIS, AND BRAIN.  
 CC -!- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF042822; AAO22230.3; -  
 CC DR EMBL; BC005496; AAH05496.1; -  
 CC DR HSSP; P20231; 1AAO.  
 CC DR MEROPS; S01\_014; -  
 CC DR MGD; MGI:133888; Scl4.  
 CC DR InterPro; IPR000859; CUB.  
 CC DR InterPro; IPR002172; LDL\_recept\_A.  
 CC DR InterPro; IPR001254; Trypsin.  
 CC DR Pfam; PF00431; CUB; 2.  
 CC DR Pfam; PF00057; ldl\_recept\_a; 4.  
 CC DR Pfam; PF00089; trypsin\_1.  
 CC DR PRINTS; PR00261; LDLRECEPTOR.  
 CC DR PRINTS; PR00722; CHYMOTRYPSIN.  
 CC DR SMART; SM00042; CUB; 2.  
 CC DR SMART; SM00192; LDLA; 3.  
 CC DR PROSITE; PS01209; LDLRA\_1; 2.  
 CC DR PROSITE; PS01209; LDLRA\_2; 4.  
 CC DR PROSITE; PS50240; TRYPsin\_DOM; 1.  
 CC DR PROSITE; PS00134; TRYPsin\_HIS; 1.  
 CC DR PROSITE; PS00135; TRYPsin\_SER; 1.  
 CC KW Signal-anchor; Glycoprotein; Hydrolase; Serine protease;  
 KW Transmembrane; Repeat.  
 FT DOMAIN 1 55  
 FT TRANSMEM 56 76  
 FT  
 FT DOMAIN 77 855  
 FT DOMAIN 214 331  
 FT DOMAIN 340 444  
 FT DOMAIN 451 488  
 FT DOMAIN 489 522  
 FT DOMAIN 523 561  
 FT DOMAIN 565 604  
 FT DOMAIN 615 854  
 FT ACT\_SITE 656 656  
 FT ACT\_SITE 711 711  
 FT ACT\_SITE 805 805  
 FT CARBOHYD 107 107  
 FT CARBOHYD 302 302  
 FT CARBOHYD 365 365  
 FT CARBOHYD 421 421  
 FT CARBOHYD 489 489  
 FT CARBOHYD 772 772  
 SQ SEQUENCE 855 AA; 94654 MW; 4F10B84DA2146DD5 CRC64;

Query Match 87.3%; Score 1152; DB 1; Length 855;  
 Best Local Similarity 86.3%; Pred. No. 3.8e-101;  
 Matches 208; Conservative 19; Mismatches 14; Indels 0; Gaps 0;

QY 1 VVGTDADGEGPQVSLHALGQGHICGASLISPNWLVSAARHYIDRGRYSPTQWTA 60  
 DB 615 VVGGINADGEGPQVSLHALGQGHICGASLISPNWLVSAARHYIDRGRYSPTQWTA 674  
 QY 61 FLGLHQSORAPGQVQRRLKRITSHPFNFDFDYDIALLEKPKREYSMPICLPD 120  
 DB 675 FLGLDQSKRSASGQVQLKRIITHPSFNDFDYDIALLEKPKREYSMPICLPD 734  
 QY 121 ASHVPFAGKAIWVTGWTGQGTGALLQKGETRVINOTTENLLPQQTIPRMVCGFL 180  
 DB 735 ATRVFPAGKAIWVTGWTGQGTGALLQKGETRVINOTTENLLPQQTIPRMVCGFL 794  
 QY 181 SGGVSCQSGDGGPLSSVVEADGRIFQAGVSWGDCGCAQRNKPQVYTRLPFRDWIKENTG 240  
 DB 795 SGGVSDQSGDGGPLSSAEKDKGRMFQAGVSWGDCGCAQRNKPQVYTRLPVYVROWIKENTG 854  
 QY 241 V 241  
 DB 855 V 855

RESULT 3  
 DES1\_HUMAN STANDARD; PRT; 422 AA.  
 AC Q9UL52;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Serine protease DES1 precursor (EC 3.4.21.-).  
 GN DES1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2109480; PubMed=11161383;  
 RA Lang J.C., Schuller D.E.;  
 RT "Differential expression of a novel serine protease homologue in  
 RT squamous cell carcinoma of the head and neck";  
 RL Br. J. Cancer 84:237-243(2001).  
 CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ACTIVATED BY  
 CC CLEAVAGE AND SECRETED (By similarity).  
 CC -!- TISSUE SPECIFICITY: Expression can only be detected in tissues  
 CC derived from the head and neck, and in skin, prostate and testis.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 CC -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF084819; AAF04328.1; -  
 CC HSSP; P00763; IDPO.  
 CC MEROPS; S01\_021; -  
 CC InterPro; IPR001314; Chymotrypsin.  
 CC InterPro; IPR000082; SEA.  
 CC InterPro; IPR001254; Trypsin.  
 CC Pfam; PF01390; SEA; FALSE\_NEG.  
 CC Pfam; PF00089; trypsin; 1.  
 CC PRINTS; PR00722; CHYMOTRYPSIN.  
 CC SMART; SM00200; SEA; 1.  
 CC SMART; SM00200; Tryp\_Spc; 1.  
 CC PROSITE; PS50024; SEA; 1.  
 CC PROSITE; PS50240; TRYPsin\_DOM; 1.  
 CC PROSITE; PS00134; TRYPsin\_HIS; 1.  
 CC PROSITE; PS00135; TRYPsin\_SER; 1.  
 KW Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;  
 KW Glycoprotein.  
 FT CHAIN 1 190 SERINE PROTEASE DES1, NON-CATALYTIC  
 FT CHAIN 191 422 SERINE PROTEASE DES1, CATALYTIC CHAIN  
 FT CHAIN 422 (POTENTIAL).  
 FT DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 19 39 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT DOMAIN 40 422 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 43 164 SEA.  
 FT DOMAIN 191 421 SERINE PROTEASE.  
 FT ACT\_SITE 231 231 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 276 276 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 372 372 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 176 296 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 216 232 BY SIMILARITY.  
 FT DISULFID 341 357 BY SIMILARITY.  
 FT DISULFID 368 397 BY SIMILARITY.  
 FT CARBOHYD 74 74 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 165 165 N-LINKED (GLCNAC... ) (POTENTIAL).







DR EMBL; Y19140; CAB65555.1; JOINED.  
 DR EMBL; Y19141; CAB65555.1; JOINED.  
 DR EMBL; Y19142; CAB65555.1; JOINED.  
 DR EMBL; Y19143; CAB65555.1; JOINED.  
 DR EMBL; AL163218; CAB90392.1; .  
 DR EMBL; AL163217; CAB90389.1; .  
 DR HSP; P00763; LDPO.  
 DR MEROPS; S01.156; .  
 DR NIM; 226200; .  
 DR InterPro; IPR000859; CUB.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001172; LDL\_recept\_A.  
 DR InterPro; IPR000998; MAM.  
 DR InterPro; IPR000082; SEA.  
 DR InterPro; IPR001190; SRCR.  
 DR InterPro; IPR001254; Trypsin.  
 DR Pfam; PF00431; CUB; 2.  
 DR Pfam; PF00057; Iol\_recept\_a; 2.  
 DR Pfam; PF00629; MAM; 1.  
 DR Pfam; PF01390; SEA; 1.  
 DR Pfam; PF00530; SRCR; 1.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00042; CUB; 2.  
 DR SMART; SM00132; LDLA; 2.  
 DR SMART; SM00137; MAM; 1.  
 DR SMART; SM00200; SEA; 1.  
 DR SMART; SM00202; SR; 1.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR PROSITE; PS01180; CUB; 2.  
 DR PROSITE; PS01209; LDLA\_1; 2.  
 DR PROSITE; PS50068; LDLA\_2; 2.  
 DR PROSITE; PS00740; MAM\_1; 1.  
 DR PROSITE; PS50060; MAM\_2; 1.  
 DR PROSITE; PS50024; SEA; 1.  
 DR PROSITE; PS00420; SRCR\_1; FALSE\_NEG.  
 DR PROSITE; PS02287; SRCR\_2; 1.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS0134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Signal-anchor; Glycoprotein; Myristate; Hydrolase.  
 KW Serine protease; Zymogen; Transmembrane; Repeat.  
 FT CHAIN 1 784  
 FT CHAIN 785 1019  
 FT DOMAIN 1 18  
 FT TRANSMEM 19 47  
 FT DOMAIN 48 1019  
 FT DOMAIN 52 169  
 FT DOMAIN 182 223  
 FT DOMAIN 225 334  
 FT DOMAIN 342 504  
 FT DOMAIN 524 634  
 FT DOMAIN 641 679  
 FT DOMAIN 678 771  
 FT DOMAIN 785 1019  
 FT ACT\_SITE 825 825  
 FT ACT\_SITE 876 876  
 FT ACT\_SITE 971 971  
 FT LIPID 2 2  
 FT DISULFID 184 197  
 FT DISULFID 191 210  
 FT DISULFID 204 221  
 FT DISULFID 643 655  
 FT DISULFID 650 668  
 FT DISULFID 662 677  
 FT DISULFID 672 896  
 FT DISULFID 810 826  
 FT DISULFID 910 977  
 FT DISULFID 941 956  
 FT DISULFID 967 995  
 FT CARBOHYD 116 116  
 FT CARBOHYD 147 147

FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 328 328 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 335 335 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 388 388 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 440 440 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 503 503 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 630 630 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 682 682 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 706 706 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 848 848 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 887 887 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 909 909 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 949 949 Q -> E (IN REF. 3).  
 FT CONFLICT 134 134 S -> P (IN REF. 3).  
 FT CONFLICT 732 732 SQCLQDSLRQCNEKS -> RRNAKNEIDALSPIILIA  
 FT CONFLICT 754 771 (IN REF. 3).  
 SQ SEQUENCE 1019 AA; 112923 MW; B6AAA245F6D4A563 CRC64;

Query Match 41.5%; Score 547.5; DB 1; Length 1019;  
 Best Local Similarity 44.7%; Pred. No. 6.8e-44;  
 Matches 106; Conservative 44; Mismatches 80; Indels 7; Gaps 4;  
 QY 1 VVGFTDADEGEWPQVSLHALGQGHGICGASLISPNWLVSAAHCYVIDDRGFRYSDPTQWTA 60  
 DB 785 IVGGSNAKEGAPWVGLY-YGRLLCASLSVSDWLVSAAHCYVG----RNLEPSKWT 839  
 QY 61 FLGLHDQSRSPAGVQERRRLKRIISHPFNFDFDYDIALLELEKPAEYSSMYRPICLPD 120  
 DB 840 ILGLHMKSLTSPQTPRLIDEIVINPHYRRRRKNDNDIAMHLEFKVNYTDYIQPICLPE 899  
 QY 121 ASHVFPAGKAIWVTGWGHTQYGGTGALILQGEIRVNOTTCENLLPQ-QITPRMVCVF 179  
 DB 900 ENQVPPPGNCSAGWTVYQGITANILQEAADVPLSNERCQQQOMPEYNITENMICAGY 959  
 QY 180 LSGGVDSQCGSGGLSSVEADGRIFQAGVYVSMGDCGAQRNKFVYVTRLPFLFRDWIK 236  
 DB 960 BEGGIDSQCGSGGLMCQE-NNRWFLAGVTSGYKCALPNRPFGYARVSRFTEWIQ 1015

RESULT 6  
 ENTK\_BOVIN  
 ID ENTK\_BOVIN STANDARD; PRT; 1035 AA.  
 AC P98072;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase).  
 GN PRS7 OR ENTK.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_taxid=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Duoenum;  
 RX MEDLINE=94329561; PubMed=8052624;  
 RA Kitamoto Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.;  
 RT "Enterokinase, the initiator of intestinal digestion, is a mosaic  
 RT protease composed of a distinctive assortment of domains";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).  
 RN [2]  
 RP SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=94043122; PubMed=8226855;  
 RA Lavallie E.R., Rehemtulla A., Racie L.A., Diblasio E.A.,  
 RA "Cloning and functional expression of a cDNA encoding the catalytic  
 RT subunit of bovine enterokinase";

J. Biol. Chem. 266:23311-23317(1993).

[3]

SEQUENCE OF 801-827.

TISSUE=Intestine;

MEDLINE=92189715; PubMed=1799406;

Light A., Janska H.;

"The amino-terminal sequence of the catalytic subunit of bovine enterokinase.";

J. Protein Chem. 10:475-480(1991).

-1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN, PROCARBOXYPEPTIDASES, AND PROELASTASES.

-1- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-Ile-7 bond in trypsinogen.

-1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.

-1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).

-1- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.

-1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS CLEAVED BY A TRYPSIN-LIKE PROTEASE.

-1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.

-1- SIMILARITY: CONTAINS 2 CUB DOMAINS.

-1- SIMILARITY: CONTAINS 1 SEA DOMAIN.

-1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.

-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

-----

EMBL; U09859; AAB40026.1; -.

EMBL; L19663; AAL16035.1; -.

PIR; A61436; A61436.

HSSP; P20231; 1A00.

MEROPS; S01.156; -.

InterPro; IPR000859; CUB.

InterPro; IPR001314; Chymotrypsin.

InterPro; IPR002172; LDL\_recept\_A.

InterPro; IPR000998; MAM.

InterPro; IPR000082; SEA.

InterPro; IPR001190; SRCR.

InterPro; IPR001254; Trypsin.

Pfam; PF00431; CUB; 2.

Pfam; PF00057; ldl\_recept\_a; 2.

Pfam; PF00629; MAM; 1.

Pfam; PF01390; SEA; 1.

Pfam; PF00530; SRCR; 1.

Pfam; PF00089; trypsin; 1.

PRINTS; PR00722; CHYMOTRYPSIN.

SMART; SM00042; CUB; 2.

SMART; SM00192; LDLA; 2.

SMART; SM00137; MAM; 1.

SMART; SM00200; SEA; 1.

SMART; SM00202; SR; 1.

SMART; SM00020; TRYP\_SPC; 1.

PROSITE; PS01180; CUB; 2.

PROSITE; PS01209; LDLA\_1; 2.

PROSITE; PS00068; LDLA\_2; 2.

PROSITE; PS00740; MAM\_1; 1.

PROSITE; PS50060; MAM\_2; 1.

PROSITE; PS50024; SEA; 1.

PROSITE; PS00420; SRCR\_1; FALSE\_NEG.

PROSITE; PS50287; SRCR\_2; 1.

PROSITE; PS50240; TRYPSIN\_DOM; 1.

PROSITE; PS00134; TRYPSIN\_HIS; 1.

-----

DR PROSITE; PS00135; TRYPSIN\_SER; 1.

KW Signal-anchor; Glycoprotein; Myristate; Hydrolase;

KW Serine protease; Zymogen; Transmembrane; Repeat; Alternative splicing.

FT CHAIN 1 800 NON-CATALYTIC CHAIN (HEAVY CHAIN).

FT CHAIN 801 1035 CATALYTIC CHAIN (LIGHT CHAIN).

FT DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).

FT DOMAIN 48 1035 EXTRACELLULAR (POTENTIAL).

FT DOMAIN 54 169 SEA.

FT DOMAIN 197 238 LDL-RECEPTOR CLASS A 1.

FT DOMAIN 240 350 CUB 1.

FT DOMAIN 358 520 MAM.

FT DOMAIN 540 650 CUB 2.

FT DOMAIN 657 695 LDL-RECEPTOR CLASS A 2.

FT DOMAIN 694 787 SRCR.

FT DOMAIN 801 1035 SERINE PROTEASE.

FT ACT\_SITE 841 841 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 892 892 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 987 987 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT LIPID 2 2 MYRISTATE (POTENTIAL).

FT DISULFID 199 212 BY SIMILARITY.

FT DISULFID 206 225 BY SIMILARITY.

FT DISULFID 219 236 BY SIMILARITY.

FT DISULFID 659 671 BY SIMILARITY.

FT DISULFID 666 684 BY SIMILARITY.

FT DISULFID 678 693 BY SIMILARITY.

FT DISULFID 788 912 INTERCHAIN (BY SIMILARITY).

FT DISULFID 826 842 BY SIMILARITY.

FT DISULFID 926 993 BY SIMILARITY.

FT DISULFID 957 972 BY SIMILARITY.

FT DISULFID 983 1011 BY SIMILARITY.

FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 147 147 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 170 170 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 194 194 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 263 263 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 264 264 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 404 404 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 456 456 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 486 486 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 519 519 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 550 550 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 646 646 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 698 698 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 722 722 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 741 741 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 762 762 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 864 864 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 903 903 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).

FT VARSPLIC 166 192 MISSING (IN SHORT ISOFORM).

FT CONFLICT 808 808 R -> Y (IN REF. 3).

SQ SEQUENCE 1035 AA; 114887 MW; E207970B08296E13 CRC64;

-----

Query Match 41.3%; Score 544.5; DB 1; Length 1035;

Best Local Similarity 43.5%; Pred. No. 1.3e-43;

Matches 103; Conservative 50; Mismatches 77; Indels 7; Gaps 4;

QY 1 VVGGTDADEGEWPFQVSLHALGGHICGASLISPNWLYSAAHCYIDDRGFYSPTOWTA 60

Db 801 IVGGSRSREGANFWVVALYFDQ-QVCGASLVSRDLVSAHCYVG----RNMEPSKWA 855

QY 61 FLGLHDQSQRSPAGVQERLKRRIISHPFNDFTFDYDIALLELEKPAEYSVMRPICLP 120

Db 856 VLGLHMASNLTSPIQETRLDQIVINPHYNKRKNNDIAMHLEKVMNYTDYIQPICLPE 915

QY 121 ASHVFPAGKAIWVTGNGHTQYGGTCLILORGEIRVINTTCENLLPQ-QITPRMVCVF 179

Db 916 ENQVFPFGRICSIAGWALTYQGSTADVLQEAADVPLLSNKKCOQOOMPENITENNVKAGY 975

```
QY 180 LSGGVDSQDGGSLSSVEADGRIFQAGVYVSGDCAQNRKQVYTRLPFRDWIK 236
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 976 EAGGVDSQDGGSLSSVEMQOE-NRWLLAGVTYFQYQALPFRGVARVPRFTWQ 1031
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 7
TMS2_MOUSE STANDARD; PRT; 490 AA.
AC 09J108: 09JTK4: 090Y82:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE transmembrane protease, serine 2 (EC 3.4.21.-) (Epitheliasin) (Plasmic
GN transmembrane protein X).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN=BALB/C;
RX MEDLINE=21104370; PubMed=11169526;
RA Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Vihko P.T.;
RT "Expression of transmembrane serine protease TMPRSS2 in mouse and
RL human tissues.";
RL J. Pathol. 193:134-140(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Han J., Kim S.;
RT "Putative transmembrane protease X";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RA STRAIN=BALB/C;
RC Jacquinet E.J., Rao N.V., Rao G.N., Hoidal J.R.;
RT "A novel mosaic serine protease, epitheliasin.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ACTIVATED BY
CC -!- TISSUE SPECIFICITY: EXPRESSED MAINLY IN PROSTATE AND KIDNEY.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF199362; AAF97867.1; -.
DR EMBL; AF243500; AAF64186.1; -.
DR EMBL; AF113596; AAF21308.1; -.
DR HSSP; P00761; IAKS.
DR MGD; MGI:1354381; Tmprss2.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL_recept_A.
DR InterPro; IPR001190; SRCR.
DR Pfam; PF00057; ldl_recept_1.
DR Pfam; PF00089; trypsin.1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa.1.
DR SMART; SM00202; SR.1.
DR SMART; SM00020; Tryp-Src; 1.
DR PROSITE; PS01209; LDLRA_1.1.
DR PROSITE; PS00068; LDLRA_2.1.
DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR PROSITE; PS0287; SRCR_2; 1.
```

```
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen.
FT CHAIN 1 253
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT CHAIN 254 490
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT DOMAIN 1 83
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT DOMAIN 84 104
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT DOMAIN 105 490
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT DOMAIN 111 149
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT DOMAIN 150 242
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT DOMAIN 254 490
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT ACT_SITE 294 294
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT ACT_SITE 343 343
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT ACT_SITE 439 439
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT SITE 253 254
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT DISULFID 76 125
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT DISULFID 119 138
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT DISULFID 132 147
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT DISULFID 171 230
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT DISULFID 184 240
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT DISULFID 243 363
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT DISULFID 279 295
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT DISULFID 408 424
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT DISULFID 435 463
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT CARBOHYD 111 111
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT CARBOHYD 212 212
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT CARBOHYD 474 474
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT CONFLICT 122 122
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT CONFLICT 178 178
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT CONFLICT 320 320
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
FT CONFLICT 474 474
FT TRANSMEMBRANE PROTEASE, SERINE 2,
FT CATALYTIC CHAIN.
SQ SEQUENCE 490 AA; 53479 MW; 07D2B03EA4D8A1A9 CRC64;

Query Match 38.78; Score 510; DB 1; Length 490;
Best Local Similarity 40.94; Pred. No. 1e-40;
Matches 97; Conservative 48; Mismatches 82; Indels 10; Gaps 5;

QY 1 VVGGTDADGEGWPQVSLHALGQGHICGASLISPNVLVSAAHGVIDDRGFYSDPTQWTA 60
      :||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 254 IVGGLNASPGDMPQVSLHVQGV-HVCGGSIITPEWIVTAHCVVEP---LSGPRWTA 308
      :||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 FLGLHDQSQRSAPGVQERRLKRIISHFFNDFTDYDIALLELEKPAEYSSMYRPLCLPD 120
      :||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 309 FAGILRQSLMEFYG--SRHQVEKVISHPNYSKTKNNDIALMKLQTPLAENDLVKPYCLPN 366
      :||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 121 ASHVFPAGKATVWTGWHGTQYGGTGALILQKGEIRVNTOTTEN--LLPQOITPRMCMVG 178
      :||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 367 PGMLLDLDQECWISGWGATYKGTSDVLNAAVPLLEPSKCNKSKYIYNLLITPAMICAG 426
      :||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 179 FLISGVDSQDGGSLSSVEADGRIFQAGVYVSGDCAQNRKQVYTRLPFRDWI 235
      :||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 427 FLQGSVDSQDGGSLVTLK-NGIWWLIGTWSGSCAKALRPGVGNVTVFTDWI 482
      :||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 8
ENTK_MOUSE STANDARD; PRT; 1069 AA.
ID ENTK_MOUSE
AC P97435;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Enterocopeptidase (EC 3.4.21.9) (Enterokinase).
DE PRS7 OR ENTK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
```

SEQUENCE FROM N.A.  
 STRAIN=C57BL/6; TISSUE=Ductenum;  
 MEDLINE=38147142; PubMed=9486188;  
 Yuan X., Zheng X., Lu D., Rubin D.C., Pung C.Y.M., Sadler J.E.;  
 "Structure of murine enterokinase (enteropeptidase) and expression in  
 small intestine during development.";  
 Am. J. Physiol. 274:G342-G349(1998).  
 -!- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC  
 PROTEOLYTIC PREENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE  
 A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN  
 TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,  
 PROCARBOXYPEPTIDASES, AND PROELASTASES (BY SIMILARITY).  
 -!- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-Ile-7 bond in  
 trypsinogen.  
 -!- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A  
 MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND (BY  
 SIMILARITY).  
 -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).  
 -!- PM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS  
 CLEAVED BY A TRYPSIN-LIKE PROTEASE (BY SIMILARITY).  
 -!- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.  
 -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.  
 -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.  
 -!- SIMILARITY: CONTAINS 1 SRCR DOMAIN.  
 -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.  
 -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 TRYPSIN FAMILY.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL: U73378; AAB37317.1; -;  
 HSP; P20231; IAAO.  
 MEROPS: S01.156; -;  
 MGD: MGI:1197523; Prs7.  
 InterPro: IPR000859; CUB.  
 InterPro: IPR001334; Chymotrypsin.  
 InterPro: IPR002172; LDL\_recept\_A.  
 InterPro: IPR000998; MAM.  
 InterPro: IPR000082; SEA.  
 InterPro: IPR001190; SRCR.  
 InterPro: IPR001254; Trypsin.  
 Pfam: PF00431; CUB; 2.  
 Pfam: PF00057; ldl\_recept\_a; 2.  
 Pfam: PF00629; MAM; 1.  
 Pfam: PF01390; SEA; 1.  
 Pfam: PF00530; SRCR; 1.  
 Pfam: PF00089; trypsin; 1.  
 PRINTS: PR00722; CHYMOTRYPSIN.  
 SMART: SM00042; CUB; 2.  
 SMART: SM00192; LDLA; 2.  
 SMART: SM00137; MAM; 1.  
 SMART: SM00200; SEA; 1.  
 SMART: SM00202; SR; 1.  
 SMART: SM00020; Tryp\_Spc; 1.  
 PROSITE: PS01180; CUB; 2.  
 PROSITE: PS01209; LDLA\_1; 2.  
 PROSITE: PS00088; LDLA\_2; 2.  
 PROSITE: PS00740; MAM\_1; 1.  
 PROSITE: PS00060; MAM\_2; 1.  
 PROSITE: PS50024; SEA; 1.  
 PROSITE: PS00420; SRCR\_1; FALSE\_NEG.  
 PROSITE: PS0287; SRCR\_2; 1.  
 PROSITE: PS50240; TRYPSIN\_DOM; 1.  
 PROSITE: PS00134; TRYPSIN\_HIS; 1.  
 PROSITE: PS00135; TRYPSIN\_SER; 1.  
 Signal-anchor; Glycoprotein; Myristate; Hydrolase;  
 Serine protease; zymogen; Transmembrane; Repeat.

PT	CHAIN	1	829	NON-CATALYTIC CHAIN (HEAVY CHAIN).
PT	CHAIN	830	1069	CATALYTIC CHAIN (LIGHT CHAIN).
PT	DOMAIN	1	18	CYTOPLASMIC (POTENTIAL).
PT	TRANSMEM	19	47	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
PT	DOMAIN	48	1069	EXTRACELLULAR (POTENTIAL).
PT	DOMAIN	52	169	SEA.
PT	DOMAIN	227	268	LDL-RECEPTOR CLASS A 1.
PT	DOMAIN	270	379	CUB 1.
PT	DOMAIN	387	549	MAM.
PT	DOMAIN	569	679	CUB 2.
PT	DOMAIN	686	724	LDL-RECEPTOR CLASS A 2.
PT	DOMAIN	723	816	SRCR.
PT	DOMAIN	830	1069	SERINE PROTEASE.
PT	ACT_SITE	874	874	CHARGE RELAY SYSTEM (BY SIMILARITY).
PT	ACT_SITE	925	925	CHARGE RELAY SYSTEM (BY SIMILARITY).
PT	ACT_SITE	1021	1021	CHARGE RELAY SYSTEM (BY SIMILARITY).
PT	LIPID	2	2	MYRISTATE (POTENTIAL).
PT	DISULFID	229	242	BY SIMILARITY.
PT	DISULFID	236	255	BY SIMILARITY.
PT	DISULFID	249	266	BY SIMILARITY.
PT	DISULFID	688	700	BY SIMILARITY.
PT	DISULFID	695	713	BY SIMILARITY.
PT	DISULFID	707	722	BY SIMILARITY.
PT	DISULFID	817	945	INTERCHAIN (BY SIMILARITY).
PT	DISULFID	859	875	BY SIMILARITY.
PT	DISULFID	959	1027	BY SIMILARITY.
PT	DISULFID	991	1006	BY SIMILARITY.
PT	DISULFID	1017	1045	BY SIMILARITY.
PT	CARBOHYD	147	147	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	197	197	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	212	212	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	373	373	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	380	380	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	433	433	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	515	515	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	579	579	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	675	675	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	727	727	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	751	751	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	770	770	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	791	791	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	897	897	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	936	936	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	999	999	N-LINKED (GLCNAC. .) (POTENTIAL).
SQ	SEQUENCE	1069	AA; 118735	MW; E62549E463743C3D CRC64;

Query Match 38.1%; Score 503; DB 1; Length 1069;

Best Local Similarity 42.1%; Pred. No. 1.2e-39;

Matches 101; Conservative 46; Mismatches 83; Indels 10; Gaps 5;

QY	1	VVGTDADGEWPOVSLHALGQGH--ICGASLISPNLVSAAHCYIDDRGRYSDPTQ	57
Db	830	IVGSDAQAGANPWWVALYHRDRSTDRLLCGASLVSAAHCYVYR---RNDPTR	885
QY	58	WTAFGLHQDSQSAFGVQERRLRIISHPFNDFTDYDIALLELEKPAEYSSMVRPIC	117
Db	886	WTAFLGLHMQSNLTSPQVRRVVDQIVINPHVDRRKRKNDIAMHLEFKVNTDVIQIC	945
QY	118	LPDASHVFPAGKAIWWTGCHTQY-CGTGALILQKEIRVINQTTCCENLLPQ-QITPRMK	175
Db	946	LPEENQIFIPGRTCSAGMGYDKINAGSTVDVLKADVPLISNEKCCQQLPEYNITESMI	1005
QY	176	CVGFLSGVDSCQDGGPLSSVEADGRIFQAGVSWGDCQACQAKNPGVYTRLPFRDWI	235
Db	1006	CAGVEEGIDSCQDGGPLMCQE-NRWFLVGVTSFGVQCCLPNHPGVYVRVSQFIEMI	1064

RESULT 9

HATT\_HUMAN

ID HATT\_HUMAN

AC O60235; STANDARD; PRT; 418 AA.

DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Airway trypsin-like protease precursor (EC 3.4.21.-).  
HA1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=98234382; PubMed=9565616;  
RA Yamaoka K., Masuda K.-I., Ogawa H., Takagi K., Umemoto N., Yamaoka S.;  
RT "Cloning and characterization of the cDNA for human airway trypsin-  
RT like protease";  
RL J. Biol. Chem. 273:11895-11901(1998).  
RN [2]  
RP SEQUENCE OF 187-206, AND CHARACTERIZATION.  
RX MEDLINE=97224034; PubMed=9070615;  
RA Yasuoka S., Ohnishi T., Kawano S., Tsuchihashi S., Ogawara M.,  
RA Masuda K.-I., Yamaoka K., Takahashi M., Sano T.;  
RT "Purification, characterization, and localization of a novel  
RT trypsin-like protease found in the human airway";  
RL Am. J. Respir. Cell Mol. Biol. 16:300-308(1997).  
CC -!- FUNCTION: May play some biological role in the host defense system  
CC on the mucous membrane independently of or in cooperation with  
CC other substances in airway mucous or bronchial secretions.  
CC -!- CATALYTIC ACTIVITY: Preferentially cleaves the C-terminal side of  
CC arginine residues at the P1 position of certain peptides, cleaving  
CC Boc-Phe-Arg-4-methylcoumaryl-7-amide most efficiently and  
CC having an optimum pH of 8.6 with this substrate.  
CC -!- ENZYME REGULATION: Strongly inhibited by diisopropyl  
CC fluorophosphate, leupeptin, antipain, aprotinin, and soybean  
CC trypsin inhibitor, but hardly inhibited by secretory leukocyte  
CC protease inhibitor at 10 microm.  
CC -!- SUBUNIT: Monomer.  
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ACTIVATED BY  
CC CLEAVAGE AND SECRETION.  
CC -!- TISSUE SPECIFICITY: Located in the cells of the submucosal serous  
CC glands of the bronchi and trachea.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AB02134; BAA28691.1; -.  
DR HSP; P00750; IRTF.  
DR MIM; 605369; -.  
DR MEROPS; S01.301; -.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR000082; SEA.  
DR InterPro; IPR001254; Trypsin.  
DR Pfam; PF01390; SEA; 1.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00200; SEA; 1.  
DR SMART; SM00020; TRYP\_SPC; 1.  
DR PROSITE; PS00024; SEA; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;  
KW glycoprotein.  
FT CHAIN 1 186 AIRWAY TRYPSIN-LIKE PROTEASE, NON-  
FT CATALYTIC CHAIN.  
FT CHAIN 187 418 AIRWAY TRYPSIN-LIKE PROTEASE, CATALYTIC

FT DOMAIN 1 20 CHAIN.  
FT TRANSMEM 21 41 CYTOPLASMIC (POTENTIAL).  
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
FT (POTENTIAL)  
FT EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 42 418 SEA.  
FT DOMAIN 44 164 SERINE PROTEASE.  
FT ACT\_SITE 187 417 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 227 227 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 272 272 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 368 368 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DISULFID 173 292 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 212 228 BY SIMILARITY.  
FT DISULFID 337 353 BY SIMILARITY.  
FT DISULFID 364 393 BY SIMILARITY.  
FT CARBOHYD 144 144 N-LINKED (GLCNAC..). (POTENTIAL).  
SQ SEQUENCE 418 AA; 46263 MW; F4BC1DB020CFBBD0 CRC64;  
  
Query Match 37.78; Score 497.5; DB 1; Length 418;  
Best Local Similarity 40.5%; Pred. No. 1.3e-39;  
Matches 100; Conservative 44; Mismatches 82; Indels 21; Gaps 8;  
  
QY 1 VVGTDADGEGPQVSLHALGQGHICGASLISPNVLVSAAHGYIDDRGFYSDPTQWTA 60  
DB 187 ILGGEAEBSKPMQVSLR-LNNAHCGGSLNNMWTAAHCFRSN-----SNPRDWIA 240  
QY 61 FLGLHDQSORAGPVQERRLRKRIISHFFNDFTDYDIALLEKPAEYSSMYRPICLPD 120  
DB 241 TSGI-----STTFPKLR-MRVNRLIHNNYSATHENDIALVRLNSVTFKDIHSCVCPA 295  
QY 121 ASHVFPAGKATWVTGHTQYGGTGALIQGEIRVINQITCENLLPQQ-----ITPRMMC 176  
DB 296 ATQNPPSTAYVTGWAQAEYAGHTVPELRGQVRIISNDVCN--APHSYNGAILSGMLC 353  
QY 177 VFLSGVDSCQSGSGPLSSVEADGR--IPQAVGWGDCGCAQRNPGVYTRLPFRDW 234  
DB 354 AGVPOGGVDACQSGSGPL--VQEDSRRLMFIIVGWSGDCGLPDPGVYTRVAYLDW 411  
QY 235 IKENTGV 241  
DB 412 IRQQIGI 418  
  
RESULT 10  
TMS2\_HUMAN STANDARD; PRT; 492 AA.  
ID TMS2\_HUMAN STANDARD; PRT; 492 AA.  
AC O15393; Q9BXX1;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Transmembrane protease, serine 2 precursor (EC 3.4.21.-).  
GN TMPRSS2 OR PRSS10.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97468144; PubMed=9325052;  
RA Paoloni-Giacobino A., Chen H., Peitsch M.C., Rossier C.,  
RA Antonarakis S.E.;  
RT "Cloning of the TMPRSS2 gene, which encodes a novel serine protease  
RT with transmembrane, LDLRA, and SRCR domains and maps to 21q22.3";  
RL Genomics 44:309-320(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21309069; PubMed=11414763;  
RA Teng D.H., Chen Y., Lian L., Ha P.C., Tavtigian S.V., Wong A.K.;  
RT "Mutation analyses of 268 candidate genes in human tumor cell lines";  
RL Genomics 74:352-364(2001).  
RN [3]  
RP SEQUENCE FROM N.A.; AND MUTAGENESIS.  
RX MEDLINE=21139112; PubMed=11245484;

RA Afar D.E.H., Vivanco J., Hubert R.S., Kuo J., Chen E., Saffran D.C.,  
RA Raitano A.B., Jakobovits A.,  
RT "Catalytic cleavage of the androgen-regulated TMPRSS2 protease results  
in its secretion by prostate and prostate cancer epithelia.";  
RL Cancer Res. 61:1686-1692(2001).  
RN [4]  
RP TISSUE SPECIFICITY.  
RX MEDLINE=21104370; PubMed=11169526;  
RA Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Vihko P.T.;  
RT "Expression of transmembrane serine protease TMPRSS2 in mouse and  
human tissues.";  
RL J. Pathol. 193:134-140(2001).  
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ACTIVATED BY  
CLEAVAGE AND SECRETED.  
CC -!- TISSUE SPECIFICITY: EXPRESSED STRONGLY IN SMALL INTESTINE. ALSO  
EXPRESSED IN PROSTATE, COLON, STOMACH, AND SALIVARY GLAND.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
TRYPSIN FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 SRCR DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U75329; AAC51784.1; --  
DR EMBL: AF123453; RAD37117.1; --  
DR EMBL: AF270487; RAK29280.1; --  
DR EMBL: P00763; IDFO.  
DR MIM: 602060; --  
DR InterPro: IPR001314; Chymotrypsin.  
DR InterPro: IPR002172; LDL\_recept\_A.  
DR InterPro: IPR001190; SRCR.  
DR InterPro: IPR001254; Trypsin.  
DR Pfam: PF00057; ldl\_recept\_a; 1.  
DR Pfam: PF00089; trypsin; 1.  
DR PRINTS: PR00722; CHYMOTRYPSIN.  
DR SMART: SM00192; LDLA; 1.  
DR SMART: SM00202; SR; 1.  
DR SMART: SM00020; TRYP\_SPE; 1.  
DR PROSITE: PS01209; LDLRA\_1; 1.  
DR PROSITE: PS00068; LDLRA\_2; 1.  
DR PROSITE: PS00420; SRCR\_1; FALSE\_NEG.  
DR PROSITE: PS0287; SRCR\_2; 1.  
DR PROSITE: PS0240; TRYPsin\_DOM; 1.  
DR PROSITE: PS00134; TRYPsin\_HIS; 1.  
DR PROSITE: PS00135; TRYPsin\_SER; 1.  
KW Hydrolase; serine protease; Transmembrane; Signal-anchor; Zymogen;  
KW Polymorphism.  
FT CHAIN 1 255 TRANSMEMBRANE PROTEASE, SERINE 2, NON-  
FT CHAIN 256 492 CATALYTIC CHAIN.  
FT CHAIN 492 TRANSMEMBRANE PROTEASE, SERINE 2,  
FT CHAIN 492 CATALYTIC CHAIN.  
FT DOMAIN 1 84 CYTOLASMIC (POTENTIAL).  
FT TRANSMEM 85 105 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
(POTENTIAL).  
FT DOMAIN 106 492 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 112 149 LDL-RECEPTOR CLASS A.  
FT DOMAIN 150 242 SRCR.  
FT DOMAIN 256 492 SERINE PROTEASE.  
FT ACT\_SITE 296 296 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 345 345 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 441 441 CHARGE RELAY SYSTEM.  
FT SITE 255 256 CLEAVAGE (POTENTIAL).  
FT DISULFID 113 126 BY SIMILARITY.  
FT DISULFID 120 139 BY SIMILARITY.  
FT DISULFID 133 148 BY SIMILARITY.  
FT DISULFID 172 231 BY SIMILARITY.  
FT DISULFID 185 241 INTERCHAIN (BY SIMILARITY).

FT DISULFID 244 365 BY SIMILARITY.  
FT DISULFID 281 297 BY SIMILARITY.  
FT DISULFID 426 426 BY SIMILARITY.  
FT DISULFID 437 465 BY SIMILARITY.  
FT CARBOHYD 213 213 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARIANT 449 449 K -> N (IN DBSNP:1056602).  
FT FTID=VAR\_011692.  
FT MUTAGEN 255 255 R->Q: LOSS OF CLEAVAGE.  
FT MUTAGEN 441 441 S->A: LOSS OF ACTIVITY.  
FT CONFLICT 160 160 M -> V (IN REF. 3).  
FT CONFLICT 242 242 I -> L (IN REF. 1).  
FT CONFLICT 329 329 E -> Q (IN REF. 1).  
FT CONFLICT 489 491 RAD -> KAN (IN REF. 1).  
FT SEQUENCE 492 AA; 53891 MW; CAB44FDI74A9076B CRC64;  
Query Match 37.7%; Score 497; DB 1; Length 492;  
Best Local Similarity 41.4%; Pred. No. 1.7e-39;  
Matches 99; Conservative 45; Mismatches 81; Indels 14; Gaps 7;  
QY 1 VVGTTDADGEMPMQVSLHALCGHICGASLISPMWLYSAACVYIDRGFRYSPTQWTA 60  
Db 256 IYGESALPGAMPQVSLH-VONVHVGSGSIITPEWITVAHCVEKP-----LNNPWHWTA 310  
QY 61 FLGLHDQS-QRSAPGVQERLKRRIISHPFENDFTDYDIALLELEKFAEYSSMYRPICLP 119  
Db 311 FAGILRQSPMEYAGVQ---VEKVISHPNVDYKTKNDIALMKLQKPLTFNDLVKPVCLP 367  
QY 120 DASHVFPAGKATWVTGWHTQYGGTGALILQKGEIRVINTQTCN--LLPQIITPRMVCV 177  
Db 368 NFGMMLQPELCWISWGATEKEKTSVINAARVLLIETQRCNSRYVYDNLITPAMICA 427  
QY 178 GFLSGVSCQDGGSGPLSSVEADGRI-FQAGVYVSWGDCGCAQRNKPQVYTRLPFRDWI 235  
Db 428 GFLQGNVDSQDGGSGPL-VTSKNINWLLIGTWSGSCAKAYRPGVYGVNVTWDWI 484  
RESULT 11  
TMS5\_MOUSE STANDARD; PRT; 455 AA.  
AC Q9ER04; Q9ER03; Q9ER02;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin).  
GN TMPRSS5.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).  
RC TISSUE=Brain;  
RA Mitsui S., Yamaguchi N.;  
RT "cDNA cloning of mouse spinesin.";  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 4).  
RC TISSUE=Brain;  
RA Mitsui S., Yamaguchi N.;  
RT "Molecular cloning of mouse type 4 spinesin.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).  
CC -!- ALTERNATIVE PRODUCTS: 4 ISOFORMS: 1, 2, 3 AND 4 (SHOWN HERE);  
CC ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
TRYPSIN FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 SRCR DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
DB EMBL; AB016229; BAB20276.1; -  
DB EMBL; AB016230; BAB20277.1; -  
DB EMBL; AB016423; BAB20278.1; -  
DB EMBL; AB041037; BAB40328.1; -  
DB MGD; MGI:1933407; Tmpc55.  
DB InterPro; IPR001314; Chymotrypsin.  
DB InterPro; IPR001190; SRCR.  
DB InterPro; IPR001254; Trypsin.  
DB Pfam; PF00089; trypsin; 1.  
DB PRINTS; PR00722; CHYMOTRYPSIN.  
DB PROSITE; PS00135; TRYPSIN\_HIS; 1.  
DB PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DB PROSITE; PS00420; SRCR\_1; FALSE\_NEG.  
DB PROSITE; PS00287; SRCR\_2; 1.  
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor;  
KW Glycoprotein; Alternative splicing.  
FT DOMAIN 1 49  
FT TRANSMEM 50 70  
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
FT (POTENTIAL).  
FT EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 71 455  
FT DOMAIN 112 207  
FT DOMAIN 218 455  
FT ACT\_SITE 258 258  
FT ACT\_SITE 308 308  
FT ACT\_SITE 405 405  
FT SITE 217 218  
FT DISULFID 135 196  
FT DISULFID 148 206  
FT DISULFID 209 328  
FT DISULFID 243 259  
FT DISULFID 374 390  
FT DISULFID 401 429  
FT CARBOHYD 163 163  
FT CARBOHYD 170 170  
FT CARBOHYD 319 319  
FT CARBOHYD 375 375  
FT VARSPPLIC 1 144  
FT VARSPPLIC 1 10  
FT VARSPPLIC 1 182  
FT VARSPPLIC 183 192  
FT CONFLICT 325 325  
SQ SEQUENCE 455 AA; 49632 MW; 5CFC31789C6899AA CRC64;  
Query Match 36.6%; Score 483; DB 1; Length 455;  
Best Local Similarity 40.6%; Pred. No. 3.4e-38;  
Matches 99; Conservative 36; Mismatches 99; Indels 10; Gaps 6;  
QY 1 VVGTDADGEWQVSHLGGGHCIGASLSPNVLVSAHCHYIDDRGFRYSDFPTQNTA 60  
DB 218 IVGGQAVASGEWQVQASV-MLGSRHTCGASVLAPHVWYTAACHMY---SFLSRSLSSRWV 273  
QY 61 FLGLHDSQSRAPGVQRRLKRIISHPFFNDFTFDIDALLEKEPAEYVSVRPICLPD 120  
DB 274 HAGL--VSHGAVRHOQTWKEIIPPLISAQNHDDVALLQLRTINFSDIVDACLPA 331  
QY 121 ASHVFPAGKAIWVGTGHTQYGGT-GALLIQKEIRVINQTCEN--LIPQOITPRMVCV 177  
DB 332 KEQYFPWPGSQCVSWSGWHTDPSTHSSDITLQDTWVPLLSHLNCSNCSMTSGALTHRLMCLA 391  
QY 178 GFLGGVDSQSGDGGFLSSVEADGRIFQAVGWSGDGCAQRNKPQVYTRLPFLFRDWLKE 237  
DB 352 GYLGRADACQDGGFLVCPSPGD-TWHLVGVVSWGRGCAEPNRPQVYAKVAEFLDWIHD 450  
QY 238 NTGV 241

DB 451 TVQV 454  
RESULT 12  
STUB\_DROME STANDARD; PRT; 786 AA.  
ID STUB\_DROME AC Q05319;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Serine proteinase stubble (EC 3.4.21.-) (stubble-stubblod protein).  
GN SB OR SB-SBD.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OREGON-R;  
RX MEDLINE=93281671; PubMed=7685111;  
RA Appel L.F., Prout M., Abu-Shumays R., Hammonds A., Garbe J.C.,  
RA Fristrom D., Fristrom J.;  
RT "the Drosophila Stubble-stubblod gene encodes an apparent transmembrane serine protease required for epithelial morphogenesis.";  
RL Proc. Natl. Acad. Sci. U.S.A. 90:4937-4941(1993).  
CC -|- FUNCTION: HORMONE DEPENDENT PROTEASE REQUIRED FOR EPITHELIAL MORPHOGENESIS. HAS A DUAL FUNCTION, DETACHES IMAGINAL DISC CELLS FORM EXTRACELLULAR MATRICES THROUGH ITS EXTRACELLULAR PROTEOLYTIC DOMAIN AND TRANSMITS AN OUTSIDE-TO-INSIDE SIGNAL TO ITS INTRACELLULAR DOMAIN TO MODIFY THE CYTOSKELETON DURING MORPHOGENESIS. MAY BE ABLE TO ACTIVATE ITSELF.  
CC -|- SUBCELLULAR LOCATION: Type II membrane protein.  
CC -|- INDUCTION: BY 20-HYDROXYECYDSONE (20HE).  
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.  
CC -|- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-24 IS THE INITIATOR.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L11451; AAA28918.1; -  
HSP; P20231; IAAO.  
DR MEROPS; S01.225; -  
DR FlyBase; FBgn0003319; Sb.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Trypsin.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_SPC; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; FALSE\_NEG.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease; Transmembrane; Glycoprotein; Zymogen;  
KW Signal-anchor.  
FT CHAIN 1 542 NON-CATALYTIC CHAIN (POTENTIAL).  
FT CHAIN 543 786 CATALYTIC CHAIN (POTENTIAL).  
FT DOMAIN 1 58 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 59 80 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).  
FT DOMAIN 81 786 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 543 786 SERINE PROTEASE.  
FT DOMAIN 267 276 POLY-SER.  
FT DOMAIN 287 298 POLY-GLN.  
FT DOMAIN 391 478 SER/THR-RICH.  
FT DOMAIN 412 422 POLY-THR.  
FT DOMAIN 471 478 POLY-SER.



CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF030065; AAB84221.1; -

RC TISSUE=Brain;

RC TISSUE=Brain;

```
RX PubMed=11741986;
RA Yamaguchi N., Okui A., Yamada T., Nakazato H., Mitsui S.;
RT "Sinesin/TWPRSS", a novel transmembrane serine protease, cloned from
RT human spinal cord.
RL J. Biol. Chem. 276:0-0(2001).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
CC -!- TISSUE SPECIFICITY: Brain-specific. Predominantly expressed in
CC neurons, in their axons, and at the synapses of motoneurons in the
CC spinal cord.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB028140; BAB20375.1; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001190; SRCR.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; FALSE_NEG.
DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR PROSITE; PS02087; SRCR_2; FALSE_NEG.
DR Hydrolase; Serine protease; transmembrane; Signal-anchor;
KW Glycoprotein.
FT DOMAIN 1 49
FT TRANSMEM 50 70
FT
FT DOMAIN 71 457
FT DOMAIN 112 207
FT DOMAIN 218 457
FT ACT_SITE 258 258
FT ACT_SITE 308 308
FT ACT_SITE 405 405
FT SITE 217 218
FT SITE 135 196
FT DISULFID 148 206
FT DISULFID 209 328
FT DISULFID 243 259
FT DISULFID 374 390
FT DISULFID 401 429
FT CARBOHYD 163 163
FT CARBOHYD 170 170
FT CARBOHYD 195 195
FT CARBOHYD 319 319
FT CARBOHYD 375 375
FT SEQUENCE 457 AA; 49574 MW; 64406AB4985A2651 CRC64;

Query Match 36.2%; Score 477; DB 1; Length 457;
Best Local Similarity 41.1%; Pred No. 1.3e-37;
Matches 99; Conservative 36; Mismatches 94; Indels 12; Gaps 7;

QY 1 YVGTDADEGEWPMQVSLHALGGHICGASLSPNWLVSAAHCYIDDRGFYSPTQWTA 60
:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 218 IVGGQSVAPGRPMQASV-ALGFRHTCGSVLAPRWVYTAHCM---HSFRLARLSRWV 273
:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 FLGLHQSQRSAPGVQERLKLTIHSFPENDFTFDYDALLELEKPAEYSSMVRPGLPD 120
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 274 HAGL--VSHSAVRPHGALVERILPHLSIAQNHEDYDALLRLQTNLFSVTGAVCLPA 331
||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 ASHVFPAGKAIWVGWGHQYGGT-GALILQKGEIRVINQITCEN--LLPQOITPRMCV 177
```

```
Db 332 KEQFPKSGKRWVSGWGHTHPSYSSDMLQDTVPLFSTQLCNSSCVYSGALTTPRLCA 391
:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 178 GFLSGGVDSQGGSGGGLSSVADGRIFQ-AGVYVWGDCGAQRNKGVTYTRPLPLFRDWIK 236
:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 392 GILGRADACGDSGGPL--VCPDGDTRWLVGVVWGACAEHFGVYAKVAEFLDWIH 449
:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 237 E 237
:
Db 450 D 450

RESULT 15
KALL_MOUSE
ID KALL_MOUSE STANDARD; PRT; 638 AA.
AC P26262;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Plasma kallikrein precursor (EC 3.4.21.34) (Plasma prekallikrein)
DE (Kininogenin) (Fletcher factor).
GN KLK3 OR PK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC STRAIN=BALB/C; TISSUE=Liver;
RX MEDLINE=91090844; PubMed=2264928;
RA Seidah N.G., Sawyer N., Hamelin J., Mion P., Beaubien G.,
RA Brachpapa L., Rochemont J., Mbikay M., Chretien M.;
RT "Mouse plasma kallikrein: cDNA structure, enzyme characterization,
RT and comparison of protein and mRNA levels among species.";
RL DNA Cell Biol. 9:737-748(1990).
CC -!- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
CC -!- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES
CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
CC -!- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M58588; AAA63393.1; -.
DR FIR; A36537; KONSPL.
DR HSSP; P20231; IAAO.
DR MEROPS; S01.212; -.
DR MGD; MGI:102849; K1k3.
DR InterPro; IPR000177; Apple.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR003014; PAN.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00024; PAN; 4.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00005; APPLEDOMAIN.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00223; APPLE; 4.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS00495; APPLE; 4.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
```

DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;  
KW Fibrinolysis; Blood coagulation; Inflammatory response; Liver;  
KW Repeat.  
FT SIGNAL 1 19 PLASMA KALLIKREIN HEAVY CHAIN.  
FT CHAIN 20 390 PLASMA KALLIKREIN LIGHT CHAIN.  
FT DOMAIN 391 638 APPLE 1.  
FT DOMAIN 110 195 APPLE 2.  
FT DOMAIN 200 285 APPLE 3.  
FT DOMAIN 291 376 APPLE 4.  
FT DOMAIN 389 621 SERINE PROTEASE.  
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT CARBOHYD 453 453 O-LINKED (PROBABLE).  
FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT ACT\_SITE 434 434 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 483 483 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 578 578 CHARGE RELAY SYSTEM.  
FT DISULFID 21 104 BY SIMILARITY.  
FT DISULFID 47 77 BY SIMILARITY.  
FT DISULFID 51 57 BY SIMILARITY.  
FT DISULFID 111 194 BY SIMILARITY.  
FT DISULFID 137 166 BY SIMILARITY.  
FT DISULFID 141 147 BY SIMILARITY.  
FT DISULFID 201 284 BY SIMILARITY.  
FT DISULFID 227 256 BY SIMILARITY.  
FT DISULFID 231 237 BY SIMILARITY.  
FT DISULFID 292 375 BY SIMILARITY.  
FT DISULFID 318 347 BY SIMILARITY.  
FT DISULFID 322 328 BY SIMILARITY.  
FT DISULFID 340 345 BY SIMILARITY.  
FT DISULFID 383 503 BY SIMILARITY.  
FT DISULFID 419 435 BY SIMILARITY.  
FT DISULFID 517 584 BY SIMILARITY.  
FT DISULFID 548 563 BY SIMILARITY.  
FT DISULFID 574 602 BY SIMILARITY.  
SQ SEQUENCE 638 AA; 71368 MW; CC27C93AB1086599 CRC64;

Query Match 36.2%; Score 477; DB 1; Length 638;  
Best Local Similarity 39.3%; Pred. No. 1.8e-37;  
Matches 96; Conservative 43; Mismatches 91; Indels 14; Gaps 7;  
QY 1 VVGTFDADGEWPMQVSLHA--LQGHICGASLISPNWLVSAAHCYIDDRGRYSPTQW 58  
DB 391 IVGTFNASLGEWPMQVSLQVSLVSTHCCGSGTIGROWLTAHCF---DGIPY--PDVW 445  
QY 59 TAFGLHDQSQ--RSAPGVQERRLKRIISHFPNDFDYDIALLELEKPAEYSSMVRPI 116  
DB 446 RIYGGILSLSEITKTP---SSRIKELIIHQEVKVSSENYDIALIKLQPLNRYTEFKPI 502  
QY 117 CLPDASHVFPAGKAIWVGHTQYGGTGALILQKGEIRVINQTTGENTLLPQQ-ITPRNM 175  
DB 503 CLPSKADNTIYTCWVTGWGTYTKEQGETONILQKATIPLVNEECQKRYRVINKQMI 562  
QY 176 CVGFLSGGVDSQCGDSSGVSSVEADGRIFQAGVSWMGDCQAQKRPYVYTRLPFRDWI 235  
DB 563 CAGYKEGEGTDACKDGGGPL-VCKHSGRWLVGITSWGECKGRKQDPGVYTKVSEYMDWI 621  
QY 236 KENT 239  
DB 622 LEKT 625

Search completed: July 22, 2002, 17:12:43  
Job time: 128 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2002, 17:11:05 ; Search time 26.51 Seconds  
(without alignments)  
1572.681 Million cell updates/sec

Title: US-09-657-986b-2  
Perfect score: 1319  
Sequence: 1 VVGGTDADEGEWPNQVSLHA.....PGVYTRLPFRDWIKENTGV 241

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL19.\*

1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mnc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1151	87.3	855	11 Q9J117	Q9j117 rattus norv
2	865	55.6	845	13 Q9DGR1	Q9dgr1 xenopus lae
3	585.5	44.4	799	11 Q9DB10	Q9db10 mus musculu
4	524.5	39.8	279	11 Q9QZ74	Q9qz74 rattus norv
5	509	38.6	767	13 Q9DGR2	Q9dgr2 xenopus lae
6	501.5	38.0	250	5 Q9V514	Q9v514 drosophila
7	500	37.9	490	11 Q920K3	Q920k3 rattus norv
8	497	37.7	492	4 Q96T73	Q96t73 homo sapien
9	489	37.1	389	13 Q9PVX7	Q9pvx7 xenopus lae
10	483	36.6	787	5 Q9VER6	Q9vey6 drosophila
11	478.5	36.3	624	11 Q9DAT3	Q9dat3 mus musculu
12	474.5	36.0	624	11 Q91Y47	Q91y47 mus musculu
13	463.5	35.1	1524	13 Q91674	Q91674 xenopus lae
14	463	35.1	643	6 Q97506	Q97506 sus scrofa
15	462	35.0	329	13 Q42272	Q42272 xenopus lae
16	461	35.0	274	5 O16133	O16133 anopheles s

#### ALIGNMENTS

```

RESULT 1
Q9J117 ID Q9J117 PRELIMINARY; PRT; 855 AA.
AC Q9J117;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE MEMBRANE BOUND SERINE PROTEASE (MEMBRANE BOUND ARGININE SPECIFIC
DE SERINE PROTEASE).
GN MSP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=JEJUNUM;
RA Tsuzuki S.;
RT "A membrane bound serine protease expressed in rat small intestine.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=DUODENUM;
RA Itoue H., Takahashi K., Kishi K.;
RT "membrane-bound arginine specific serine protease.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
DR EMBL; AB037898; BAB03502.1; -
DR EMBL; AB049189; BAB13765.1; -
DR HSSP; P00763; IDPO.
DR MEROPS; S01.302; -
DR InterPro; IPR002106; AA_rnase_ligase_II.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR002172; LDL_recept_A.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00431; CUB; 2.

```

```

17 461 35.0 537 4 Q9BYE1 Q9byel homo sapien
18 461 35.0 581 4 Q9BYE2 Q9bye2 homo sapien
19 455.5 34.5 812 11 Q91WJ5 Q91wj5 mus musculu
20 453 34.3 258 5 Q97399 Q97399 phaeton coc
21 453 34.3 334 6 Q46507 Q46507 papio hamad
22 453 34.3 810 4 Q15146 Q15146 homo sapien
23 449 34.0 624 6 Q35WE7 Q35me7 oryctolagus
24 448.5 34.0 263 5 O02570 O02570 culex quinq
25 447.5 33.9 263 5 O62562 O62562 penaeus van
26 446 33.8 339 11 Q99L44 Q99l44 mus musculu
27 445 33.7 274 5 Q17086 Q17086 anopheles s
28 444.5 33.7 812 11 Q9R0W3 Q9r0w3 rattus norv
29 444 33.7 454 6 Q46506 Q46506 papio hamad
30 442.5 33.5 317 13 Q9DGR3 Q9dgr3 xenopus lae
31 442 33.5 261 13 Q9W7Q4 Q9w7q4 paralichthy
32 442 33.5 273 6 Q9XSM1 Q9xsm1 ovis aries
33 442 33.5 364 5 Q917V4 Q917v4 drosophila
34 436.5 33.1 467 5 Q967X8 Q967x8 panulirus a
35 435.5 33.0 297 11 Q88781 Q88781 rattus ratt
36 435 33.0 263 11 Q9CR35 Q9cr35 mus musculu
37 433.5 32.9 461 6 Q95ND7 Q95nd7 pan troglod
38 433 32.8 242 13 Q9W7Q7 Q9w7q7 paralichthy
39 433 32.8 405 4 Q96E86 Q96e86 homo sapien
40 432 32.8 263 11 Q9DC86 Q9dc86 mus musculu
41 431 32.7 273 11 Q921N4 Q921n4 mus musculu
42 429.5 32.6 310 11 Q9QYZ9 Q9qyz9 mus musculu
43 429.5 32.6 456 4 Q14316 Q14316 homo sapien
44 429 32.5 242 13 Q92099 Q92099 paranoctothe
45 429 32.5 263 11 Q9D8X8 Q9d8x8 mus musculu

```

DR	PF00057; ldl_recept_a; 4.
DR	pfam; PF00089; trypsin; 1.
DR	PRINTS; PR00722; CHYMOTRYPSIN.
DR	PRINTS; PR00261; LDLRECEPTOR.
DR	SMART; SM00042; CUB; 2.
DR	SMART; SM00192; LDLa; 3.
DR	SMART; SM00020; Tryp_SPC; 1.
DR	PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR	PROSITE; PS01180; CUB; 2.
DR	PROSITE; PS01209; LDLRA_1; 2.
DR	PROSITE; PS50068; LDLRA_2; 4.
DR	PROSITE; PS50240; TRYPSIN_DOM; 1.
DR	PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR	PROSITE; PS00135; TRYPSIN_SER; 1.
KW	Glycoprotein; Hydrolase; Protease; Serine protease.
FT	665
VARIANT	K -> N.
SEQUENCE	855 AA; 94955 MW; 35806B7ECF6CF03D CRC64;

Query Match	87.3%	Score 1151;	DB 11;	Length 855;
Best Local Similarity	87.1%;	Pred. No. 3e-107;		
Matches 210;	Conservative 15;	Mismatches 16;	Indels 0;	Gaps 0;
y	1	VYGGTDADGEWPMQVSLHALGQGHICGASLI	SPNWLVSAAHCYIDDRGFYSDP	TQWTA 60
b	615	VYGGTNADEGWPQVSLHALGQGHLCGASLI	SPWLVSAAHCFQDETIFKYS	SDHTMTTA 674
y	61	FLGLHDQSRAPGVQRRLKRIISHPFN	DDTFDYDIALLEKPAEYSSMVR	PICLPD 120
b	675	FLGLLDQKSASGVQSHKLRITHP	SNDTFDYDIALLEKPAEYSTVVR	PICLPD 734
y	121	ASHVFPAGKAIWTVGWGHTGYG	TGTHALILQKEIRVINOTTCE	NLLPQQTPRMCMVGFL 180
b	735	NTHVFPAGKAIWTVGWGHTKEG	TGTHALILQKEIRVINOTTCE	NLLPQQTPRMCMVGFL 794
y	181	SGGVDSGCGSGGPLSSVEADGR	IFQAGVYVNGDCAORNKPGVY	TRLPLFRWIKENTG 240
b	795	SGGVDSGCGSGGPLSSVEADGR	IFQAGVYVNGDCAORNKPGVY	TRLPEVRDWIKEQTG 854
y	241	V	241	
b	855	V	855	

```

RESULT      2
Q9DGR1
ID          PRELIMINARY;      PRT;      845 AA.
AC          Q9DGR1;
DT          01-MAR-2001 (TRENBLrel. 16, Created)
DT          01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT          01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE          DE HOMOLOG OF HUMAN MT-SPL.
GN          XMT-SPL.
OS          Xenopus laevis (African clawed frog).
OC          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC          Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC          Xenopodinae; Xenopus.
OX          NCBI_TaxId=8355;
RN          [1]
RP          SEQUENCE FROM N.A.
RX          MEDLINE=20363741; PubMed=10903452;
RA          Yamada K., Takabatake T., Takeshima K.;
RT          "Isolation and characterization of three novel serine protease genes
RT          from Xenopus laevis.";
RL          Gene 252:209-216(2000).
RC          -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC          TRYPsin FAMILY.
CC          -|- SIMILARITY: CONTAINS 2 CUB DOMAINS.
DR          EMBL; AB038498; BAB08218.1; -.
DR          HSSP; P00763; IDPO.
DR          MEROPS; S01.050; -.
DR          InterPro; IPR001314; Chymotrypsin.
DR          InterPro; IPR000859; CUB.

```

```

DR      InterPro: IPR002172; LDL_recept_A.
DR      InterPro: IPR001254; tryPsin.
DR      Pfam: PF00431; CUB; 2.
DR      Pfam: PF00057; ldl_recept_a; 4.
DR      Pfam: PF00089; tryPsin; 1.
DR      PRINTS: PR00722; CHYMOTRYPSIN.
DR      PRINTS: PR00261; LDLRECEPTOR.
DR      SMART: SM000042; CUB; 2.
DR      SMART: SM00192; LDLa; 3.
DR      SMART: SM00020; Tryp_SPC; 1.
DR      PROSITE: PS01180; CUB; 2.
DR      PROSITE: PS00068; LDLRA_2; 4.
DR      PROSITE: PS02040; TRYPSIN_DOM; 1.
DR      PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR      glycoprotein; Hydrolyase; Serine protease.
DR      SW
DR      SEQUENCE      845 aa; 93597 MW; 7FD7E62851A758B7 CRC64;

```

Query Match	65.6%	Score	865	DB	13	Length	845
Best Local Similarity	62.2%	Pred. No.	2.1e-78				
Matches	150	Conservative	41	Mismatches	50	Indels	0
Gaps							
Qy	1	VWGTDADEGEWPQVSLHALQGHICGASLISPNWLVSAAHCYIDDRGFYSDDPTQWTA	50				
Db	605	IVGVNADTGEFPQVSLHAKNKKHTCCASLGFFTMLISAAHCQDDHQMYSDAISLWTA	564				
Qy	61	FLGLHQSQRSAPQVGERLKRISHPFENFTFYDIALLELEKPAEYSSMVRPTCLPD	120				
Db	665	YLGLHQAQLNTKVVVEREIKRIMAHIGFNONTYDNDIAVLELEKPEYTDFTQPCVPE	724				
Qy	121	ASHVFPAGRAIWTGWGHTQYGGTALILQKGEIRVINQTTCENLPLQQTPPRMVCVGL	180				
Db	725	STHDFPVGKPIWWTGWALKKEGGAAVILQKAETIRINQTECNLLDGLQTPRLMCAGFV	784				
Qy	181	SGGVDCQDSDGGGLSSVREADGRITQAGVSWGDCGAORNGPYVYTRLPFRFROWIKENTG	240				
Db	785	SGGLDACQDSDGGGLSSVELNKNVILAGVWSWGEGCARRNKPGYTTKVSMMRDWSKDKTG	844				
Qy	241	V	241				
Db	845	L	845				

```

RESULT 3
Q9DB10 PRELIMINARY; PERT; 799 AA.
ID Q9DB10
AC Q9DB10;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE D1300008A22RIK PROTEIN.
GN D1300008A22RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER.
RX MEDLINE=21085560; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukuishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamakata I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaldo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Anco H., Balderelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

```

```
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL; AK004939; BAB23684.1; -.
DR HSSP; P00763; IDPO.
DR MGD; MGI:1919003; 1300008A22Rik.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR002172; LDL_recept_A.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00057; ldl_recept_a; 3.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00042; CUB; 1.
DR SMART; SM00192; LDLa; 3.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS01180; CUB; 1.
DR PROSITE; PS00069; LDLRA_2; 3.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN.1.
KW Glycoprotein; Hydrolase; Serine protease.
SQ SEQUENCE 799 AA; 89557 MW; 16315A646A4D5288 CRC64;

Query Match 44.4%; Score 585.5; DB 11; Length 799;
Best Local Similarity 45.6%; Pred. No. 2.9e-50;
Matches 108; Conservative 43; Mismatches 81; Indels 5; Gaps 3;

QY 1 VVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGRYSDPTQWT 60
DQ 565 IVGGTVSESGEWPQASLQIRGR-HICGGLIADRVRVITAAHCFQED---SMASPKLTV 620
QY 61 FLGLHDQSQRSAPGVQERRLRKRIIHPFFNDFTFDYDIALLELEKPAEYSSMVRPICLPD 120
DQ 621 FLGKMRQNSR-WPGEVSEKVSRLFLHPYHEEDSHDYDVALQLDHPVYSATVRPVCLFA 679
QY 121 ASHVPFPAKAIWVTGWGHTQYGGTCALILQKGEIRVINTQTCENLLPQITPRMCMVGLF 180
DQ 680 RSHFFEPQCHWITGWAQREGVSVNTLQKVDVQLVQDLCEAYRYQVSPRLCAGYR 739
QY 181 SGGVDSGCGSGPLSSVEADGRIFQAGVYSGWGDCAQRNKPQVYTRLPFLRDIWE 237
DQ 740 KKKKDACGDSGGLVCEPSGRWFLAGLYSWGLGCRPNFEGVYTRVTRVINWIIQ 796

RESULT 4
Q9Q274 PRELIMINARY; PRT; 279 AA.
AC Q9Q274;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ADRENAL SECRETORY SERINE PROTEASE PRECURSOR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEDE;
RX MEDLINE=21332587; PubMed=11439186;
RA Bicknell A.B., Lomthaisong K., Woods R.J., Hutchinson E.G.,
RA Bennett H.P.J., Gladwell R.T., Lowry P.J.;
RT "Characterization of a Serine Protease that Cleaves Pro-gamma-
RT Melanotropin at the Adrenal to Stimulate Growth.";
```

```
RL Cell 105:903-912(2001).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AF198087; AAF13253.1; -.
DR HSSP; P00760; IAQ7.
DR MEROPS; S01.047; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN.1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease; Signal.
RT SIGNAL 1.
SQ SEQUENCE 279 AA; 30522 MW; 1C28069DF0064546 CRC64;

Query Match 39.8%; Score 524.5; DB 11; Length 279;
Best Local Similarity 41.5%; Pred. No. 1e-44;
Matches 102; Conservative 42; Mismatches 83; Indels 19; Gaps 6;

QY 1 VVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGRYSDPTQWT 59
DQ 48 IIGGTAETGDWPMQVSLQ-LNNVHHCGGTLLSNLWLTAAHC-----FRSYSNPQWT 100
QY 60 AFLGLHDQSQRSAPGVQERRLRKRIIHPFFNDFTFDYDIALLELEKPAEYSSMVRPICLP 119
DQ 101 ATFGVSTISPR-----LVRVRALLAHAEVNSITRDNDIAVQVLDPRVTFTRNIHRVCLP 155
QY 120 DASHVFPAGKAIWVTGWGHTQYGGTCALILQKGEIRVINTQTCENLLPQ----QITPRM 175
DQ 156 AATQNIIPDSVAYATGWGSLTYGNTVTNLQGEVIRVSEVCNE--PAGYGGSVLPGLM 213
QY 176 CVGFLSGVDSGCGSGPLSSVEADGRIFQAGVYSGWGDCAQRNKPQVYTRLPFLRDI 235
DQ 214 CAGVRSAGVACQCGSGPLVQEDTRRLWFVVGIVSWGIVGCGLPNRPQVYTRVTRV 273
QY 236 KENTGV 241
DQ 274 RQQTGI 279

RESULT 5
Q9DGR2 PRELIMINARY; PRT; 767 AA.
AC Q9DGR2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE EMBRYONIC SERINE PROTEASE-2.
GN XESP-2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20363741; PubMed=10903452;
RA Yamada K., Takabatake T., Takeshima K.;
RT "Isolation and characterization of three novel serine protease genes
RT from Xenopus laevis.";
RL Gene 252:209-216(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AB038497; BAB08217.1; -.
DR HSSP; P00766; ICHG.
DR MEROPS; S01.049; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL_recept_A.
DR InterPro; IPR001190; SRCR.
```



OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Tsuzuki S.;  
RT "TMRSS2, Rat.";  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB073550; BAB70683.1; -;  
SQ SEQUENCE 490 AA; 53518 MW; 2BC691551CAC409A CRC64;

Query Match 37.9%; Score 500; DB 11; Length 490;  
Best Local Similarity 40.1%; Pred. No. 6.4e-42;  
Matches 95; Conservative 48; Mismatches 84; Indels 10; Gaps 5;

QY 1 VVGTDADGEGWQVSLHALGOGHICGASLISPNMLVSAACHYIDDRGFYSDPTQWTA 60  
DB 254 IVGGSTASPDGWPQVSLHVQGI-HVCGGSIITPEWIVTAREHVEEP-----LSSPRYWTA 308  
QY 61 FLGLHDQSQRSAFQVQERLKRRIISHPFNDFTFDYDIALLELEKPAEYSSMVRPCLPD 120  
DB 309 FAGILKQSLMFYG--SRHQVEKVISHPNVDSKTKNNDIALMKLQTLPLAFNDVQVPCLPN 366  
QY 121 ASHVFPAGKAIWVGHTQYGTGALILQKGEIRVINQTCEN--LLPQOITPRMVCVG 178  
DB 367 PGMLDLAQECWISGKATYKGTSDVLNAAMVPLIEPSKNSKYIYNLLITPAMICAG 426  
QY 179 FLSCGVDSQCGSGPLSSVEADGRIFQAGVSWGDCGAQRNKPQVYTRLPFRDWI 235  
DB 427 FLOGVDSQCGSGPLVTLKNE-IWNLIGDT-SWGGSCAKAYRPGVYGNVTFDWDI 482

RESULT 8  
Q96T73 PRELIMINARY; PRT; 492 AA.  
AC Q96T73; 2001 (TReMBLrel. 19, Created)  
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
DE EPIHELASIN.  
GN TMRSS2  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21233025; PubMed=11322890;  
RA Jacquinet E., Rao N.V., Rao G.V., Wang Z., Albertine K.H.,  
RA Hoidal J.R.;  
RT "Cloning and characterization of the cDNA and gene for human  
epithelialin.";  
RL Eur. J. Biochem. 268:2687-2699(2001).  
DR EMBL; AF329454; AAK53559.1; -;  
SQ SEQUENCE 492 AA; 53863 MW; 3ABA755BF276DADF CRC64;

Query Match 37.7%; Score 497; DB 4; Length 492;  
Best Local Similarity 41.4%; Pred. No. 1.3e-41;  
Matches 99; Conservative 45; Mismatches 81; Indels 14; Gaps 7;

QY 1 VVGTDADGEGWQVSLHALGOGHICGASLISPNMLVSAACHYIDDRGFYSDPTQWTA 60  
DB 256 IVGGESALPGAWPQVSLH-VQNVHVGCGSIITPEWIVTAREHVEEP-----LNNPWHWTA 310  
QY 61 FLGLHDQS-QRSAPGQVQERLKRRIISHPFNDFTFDYDIALLELEKPAEYSSMVRPCLP 119  
DB 311 FAGILRQSEFYAGYQO--VEKVISHPNVDSKTKNNDIALMKLQPLFNDLVKPCLP 367  
QY 120 DASHVFPAGKAIWVGHTQYGTGALILQKGEIRVINQTCEN--LLPQOITPRMVCV 177  
DB 368 NPGMMLQPEQLWISGWGATEERKGTSEVLNNAKVKILLIETQRNSRYVDNLLITPAMICA 427

QY 178 GFTSGGVDSQCGSGPLSSVEADGRIFQAGVSWGDCGAQRNKPQVYTRLPFRDWI 235  
DB 428 GFLQGNVDSQCGSGGPL--VTSKNNIWWLIGDT-SWGGSCAKAYRPGVYGNVTFDWDI 484  
RESULT 9  
Q9PVX7 PRELIMINARY; PRT; 389 AA.  
AC Q9PVX7;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE EPIDERMIS SPECIFIC SERINE PROTEASE.  
GN XEPSIN.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yamada K.;  
RT "The expression control of xepsin by non-axial and planar  
posteriorizing signals in Xenopus epidermis.";  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
TRYPSIN FAMILY.

DR EMBL; AS018694; BAA84941.1; -;  
DR HSSP; P00763; IDPO.  
DR MEROPS; S01.00A; -;  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Trypsin.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; TRYP-SPC; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Protease; Serine protease.  
SQ SEQUENCE 389 AA; 42375 MW; B31FB4A2F5D1F6E3 CRC64;

Query Match 37.1%; Score 489; DB 13; Length 389;  
Best Local Similarity 40.6%; Pred. No. 6e-41;  
Matches 102; Conservative 39; Mismatches 86; Indels 24; Gaps 8;

QY 1 VVGTDADGEGWQVSLHALGOGHICGASLISPNMLVSAACHYIDDRGFYSDPTQWTA 60  
DB 26 IVGMDSKRGENPQVSL-SYKSDSICGSLTDSWVTAHC-IDS-----LDVSYIV 78  
QY 61 FLGLHDQSQRSAF--GVQERLKRRIISHPFNDFTFDYDIALLELEKPAEYSSMVRPCL 118  
DB 79 YLGAY---QLSAPDNSTVSRGVKSIITKHPDFQVSGSGDIALIELEKPVTFPYILPICL 135  
QY 119 PDASHVFPAGKAIWVGHTQYGTG--TGALILQKGEIRVINQTCENLLPQO----- 169  
DB 136 PSQDVQPAAGTMCWV-TGWNIGQETPLISKTIQKAEVAIIDSSVCGTWYESSLYIDPF 195  
QY 170 --ITPRMVCVFLSGGVDSQCGSGPLSSVEADGRIFQAGVSWGDCGAQRNKPQVYTR 227  
DB 196 SFIQEDMVCAGYKGRIDACQSGSGPL-VCNVNWNVLQGLIVSWGCGCAEPNRPVTVTK 254  
QY 228 LPLFRDWIKEN 238  
DB 255 VQYVQDWLKTN 265

RESULT 10  
Q9VEY6 PRELIMINARY; PRT; 787 AA.  
ID Q9VEY6  
AC Q9VEY6; 2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)



```

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE SB GENE PRODUCT.
GN SB OR CG4316.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20195006; PubMed=10731132;
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Cocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Daventport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Murzyn D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pachter J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Zhu S., Zhu X., Smith H.O.,
RA "The genome sequence of Drosophila melanogaster."
RT Science 287:2185-2195(2000).
RL CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY.
CC EMBL; AF003712; AAF55277.1; -.
CC HSSP; P00763; LDPO.
CC MEROPS; S01.225; -.
CC FlyBase; FBgn0003319; SB.
CC InterPro; IPR001314; Chymotrypsin.
CC InterPro; IPR002965; P_Rich_extensn.
CC InterPro; IPR001254; Trypsin.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPsin.
CC PRINTS; PR01217; PRICHEXTENSIN.
CC SMART; SM00020; Tryp_SPC; 1.
CC PROSITE; PS02040; TRYPsin_DOM; 1.
CC PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 787 AA; 85143 MW; 02B286750BD6025A CRC64;

Query Match 36.6%; Score 483; DB 5; Length 787;
Best Local Similarity 39.2%; Pred. No. 6,1e-40;
Matches 98; Conservative 47; Mismatches 85; Indels 20; Gaps 8;

```

— — — — —

Db 756 PSSKEDACQSGDGLVPCONEKEQFSIYGLVSWGEGCGRVSKPGVYTKVRLFFTWI-QN 814

QY 239 T 239

Db 815 T 815

RESULT 14

O97506 PRELIMINARY; PRT; 643 AA.

AC O97506; 01-MAY-1999 (Tremblrel. 10, Created)

DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE KALLIKREIN

OS Sus scrofa (Pig)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

OX NCBI\_TaxID=9823;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=PLASMA;

RX MEDLINE=20413446; PubMed=10954859;

RA Kimura A., Kihara T., Okimura H., Hamabata T., Ohnishi J.,

RA Moriyama A., Takahashi K., Takahashi T.; plasma kallikrein, and its

RT "Identification of porcine follipisin as plasma kallikrein, and its

RT possible involvement in the production of bradykinin within the

RT follicles of porcine ovaries";

RL Mol. Reprod. Dev. 57:79-87(2000).

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

CC TRYPSIN FAMILY.

DR EMBL; AB022425; BAA37147.1; -.

DR HSSP; P00786; ICHG.

DR MEROPS; S01.212; -.

DR InterPro; IPR000177; Apple.

DR InterPro; IPR001314; Chymotrypsin.

DR InterPro; IPR003014; PAN.

DR InterPro; IPR001254; Trypsin.

DR Pfam; PR00024; PAN; 4.

DR Pfam; PR00089; trypsin; 1.

DR PRINTS; PR00005; APPLEDOMAIN.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR SMART; SM00223; APPLE; 4.

DR SMART; SM00020; TRYP\_SPE; 1.

DR PROSITE; PS00495; APPLE; 4.

DR PROSITE; PS0240; TRYPSIN\_DOM; 1.

DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.

DR PROSITE; PS00135; TRYPSIN\_SER; UNKNOWN\_1.

KW Hydrolase; Serine protease.

SQ SEQUENCE 643 AA; 72227 MW; AFE2923E3C3CB80A CRC64;

Query Match 35.1%; Score 463; DB 6; Length 643;

Best Local Similarity 40.2%; Pred. No. 4.9e-38;

Matches 98; Conservative 41; Mismatches 91; Indels 14; Gaps 6;

QY 1 VVGSTDADEGEWPQVSLHA--LGQGHICGASLISPNWLSAAHCYIDDRGFYSDPTQW 58

Db 399 IVGSTDSEFLGEPWQVSLQAKLRAQNLHCGSGSIQHWLTAACHFDG-----LSLPDIW 453

QY 59 TAFGLHDDQSQ--RSAPGVQVERLKRISHPFNDFTFDYDIALLEKPAEYSNWRPI 116

Db 454 RIYGGILNISEITKETPFSQ---VKEIIHQNYKILESCHDIALKLEPLNYDFQKPI 510

QY 117 CLPDASHVFPAGKAIWYTGHTQYGTGALILQKEIRVINGTTCCE-NLLPQOITPRM 175

Db 511 CLPRDDTNVYVNCWVTGNGFTKEGEIQNIIQKYNIPLVSENECKSYRDHKISKQMI 570

QY 176 CVGFLSGVDSQDGGSLVSEADGRIFQAGVSWGDCGCAQRNKPQVYTLPLFRDWI 235

Db 571 CAGYKEGKDGACKGESGGL-VCKYINGIHLVGTTSWEGCGARREQPGVYTKVIEYMDWI 629

QY 236 KENT 239

Db 630 LEKT 633

RESULT 15

O42272 PRELIMINARY; PRT; 329 AA.

AC O42272; 01-JAN-1998 (Tremblrel. 05, Created)

DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE SERINE PROTEASE.

GN CAP1.

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

OC Xenopodinae; Xenopus.

OX NCBI\_TaxID=8355;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=EPITHELIUM;

RX MEDLINE=97474477; PubMed=9335501;

RA Vallet V., Chraïbi A., Gaeggeler H.P., Horisberger J.D., Rossier B.C.;

RA "An epithelial serine protease activates the amiloride-sensitive

RT sodium channel";

RL Nature 389:607-610(1997).

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

CC TRYPSIN FAMILY.

DR EMBL; AF029404; AAB96905.1; -.

DR HSSP; P00763; LDPO.

DR MEROPS; S01.158; -.

DR InterPro; IPR001314; Chymotrypsin.

DR InterPro; IPR001254; Trypsin.

DR Pfam; PR00089; trypsin; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR SMART; SM00020; TRYP\_SPE; 1.

DR PROSITE; PS0240; TRYPSIN\_DOM; 1.

DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.

DR PROSITE; PS00135; TRYPSIN\_SER; 1.

KW Hydrolase; Protease; Serine protease.

SQ SEQUENCE 329 AA; 35276 MW; 71D1D56EB09A8E0 CRC64;

Query Match 35.0%; Score 462; DB 13; Length 329;

Best Local Similarity 41.5%; Pred. No. 2.6e-38;

Matches 103; Conservative 41; Mismatches 82; Indels 22; Gaps 7;

QY 1 VVGSTDADEGEWPQVSLHALGQGHICGASLISPNWLSAAHCYIDDRGFYSDPTQWTA 60

Db 30 IVGENATPKKFPQVSLRYNGR-HVCGASLISNIIITAAHCPSDH--LMSYKVIYLG 86

QY 61 FLGLHDQSQRSAPGVQVERLKRISHPFNDFTFDYDIALLEKPAEYSNWRPICLPD 120

Db 87 VLQLEYPTSES---QLLSLKEIIHPSYSHDSTGTGVALAALDPPATFSNVQPIPLPD 142

QY 121 ASHVFPAGKAIWYTGHTQYGT--GALLQKGEIRVINGTTCENLLPQOITPR----- 173

Db 143 ENVOFFIGMNCQVYTGNGTQQGVSLPGSKTQVGNVKNKIIISROTNCNL--YHINFSSSLG 200

QY 174 -----MMCVGFLSGVDSQDGGSLVSEADGRIFQAGVSWGDCGCAQRNKPQVYTL 228

Db 201 SVQDDMLCAGSAGSYDACQDGGGLTCT-VNNOPYLAADVSWGDECGAPNRPVYILI 259

QY 229 PLFRDWIK 236

Db 260 SLYSSWIR 267

Search completed: July 22, 2002, 17:13:17

Job time: 132 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2002, 17:08:39 ; Search time 13.06 Seconds  
(without alignments)  
450.733 Million cell updates/sec

Title: US-09-657-986B-2

Perfect score: 1319  
Sequence: 1 VVGSTDADEGEWQVSLHA.....PGVYTRLPFRDWIKENTGV 241

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/2/iaa/5A-COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/iaa/5B-COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/iaa/5A-COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/iaa/5B-COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/iaa/5A-COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/iaa/5B-COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1319	100.0	855	2	US-09-027-337-2
2	587.5	41.5	235	4	US-08-944-483-65
3	544.5	41.3	235	3	US-08-807-151-3
4	544.5	41.3	235	4	US-09-478-957-3
5	544.5	41.3	235	1	US-08-200-900A-2
6	544.5	41.3	798	5	PCT-US94-00616-2
7	497.5	37.7	232	1	US-08-508-448C-19
8	497.5	37.7	418	1	US-08-508-448C-25
9	497	37.7	482	4	US-09-342-749-2
10	488	37.0	283	3	US-08-807-151-1
11	488	37.0	283	4	US-09-478-957-1
12	485	36.8	454	4	US-09-518-046-2
13	481.5	36.5	415	2	US-09-000-846-2
14	473.5	35.9	225	4	US-08-944-483-67
15	473.5	35.9	256	2	US-09-027-337-3
16	469	35.6	638	2	US-08-681-151-3
17	462	35.0	248	4	US-08-944-483-63
18	461	35.0	299	4	US-08-944-483-66
19	455	34.5	238	4	US-08-944-483-64
20	453.5	34.4	812	1	US-08-248-629A-1
21	453.5	34.4	812	1	US-08-451-932-1
22	453.5	34.4	812	1	US-08-452-260-1
23	453.5	34.4	812	2	US-08-326-785-1
24	453.5	34.4	812	2	US-08-612-788-1
25	453.5	34.4	812	2	US-08-598B-1
26	453.5	34.4	812	2	US-08-429-743-1
27	453.5	34.4	812	2	US-08-866-735-1

28 453.5 34.4 812 3 US-09-066-028-1 Sequence 1, Appli  
29 453.5 34.4 812 5 PCT-US95-05107-1 Sequence 1, Appli  
30 453 34.3 791 1 US-08-843-219-1 Sequence 1, Appli  
31 453 34.3 791 3 US-08-851-350-1 Sequence 1, Appli  
32 452 34.3 230 1 US-08-456-840-47 Sequence 47, Appli  
33 452 34.3 230 1 US-08-266-407A-47 Sequence 47, Appli  
34 452 34.3 230 2 US-08-892-544-47 Sequence 47, Appli  
35 452 34.3 230 4 US-08-766-982-12 Sequence 12, Appli  
36 452 34.3 230 4 US-08-944-483-53 Sequence 12, Appli  
37 452 34.3 230 4 US-09-296-219-12 Sequence 12, Appli  
38 452 34.3 546 6 5200340-6 Patent No. 5200340  
39 452 34.3 790 1 US-08-469-486-54 Sequence 54, Appli  
40 452 34.3 790 2 US-08-469-658-54 Sequence 54, Appli  
41 452 34.3 791 2 US-09-131-995-1 Sequence 1, Appli  
42 452 34.3 791 2 US-08-832-087B-1 Sequence 1, Appli  
43 452 34.3 791 4 US-09-132-154-1 Sequence 1, Appli  
44 452 34.3 810 1 US-07-854-603-2 Sequence 2, Appli  
45 452 34.3 810 1 US-08-147-000B-29 Sequence 29, Appli

#### ALIGNMENTS

##### RESULT 1

US-09-027-337-2

; Sequence 2, Application US/09027337B

; Patent No. 5972616

; GENERAL INFORMATION:

; APPLICANT: O'Brien, Timothy J.

; APPLICANT: Tanimoto, Hiroto

; TITLE OF INVENTION: TAGD-15: An Extracellular Serine Protease Overexpressed in

; TITLE OF INVENTION: Breast and Ovarian Carcinomas

; FILE REFERENCE: D6064

; CURRENT APPLICATION NUMBER: US/09/027,337B

; CURRENT FILING DATE: 1998-02-20

; NUMBER OF SEQ ID NOS: 13

; SEQ ID NO 2

; LENGTH: 855

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: Amino acid sequence of TAGD-15 encoded by nucleotides

; OTHER INFORMATION: 23 to 2589 of Sequence 1

; Patent No. 5972616

US-09-027-337-2

Query Match 100.0%; Score 1319; DB 2; Length 855;  
Best Local Similarity 100.0%; Pred. No. 1.5e-130;  
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVGSTDADEGEWQVSLHAGOGHICGASLISPNWLVSAAHCYIDDRGFYSDEPTQWTA 60  
Db 615 VVGSTDADEGEWQVSLHAGOGHICGASLISPNWLVSAAHCYIDDRGFYSDEPTQWTA 674  
QY 61 FLGLHDQSQRSAFGVQERRLKRIISHPFNDFDYDIALLEKPAEYSSMVRPICLPD 120  
Db 675 FLGLHDQSQRSAFGVQERRLKRIISHPFNDFDYDIALLEKPAEYSSMVRPICLPD 734  
QY 121 ASHVFPAGKAIWTGCHTQYGGTGALLIQLKGEIRVINTTCENLLPQOITPRMVCVGL 180  
Db 735 ASHVFPAGKAIWTGCHTQYGGTGALLIQLKGEIRVINTTCENLLPQOITPRMVCVGL 794  
QY 181 SGGVDSQCGSGGGLSSVEADGRIFQAGVYSGDGCQAQRNKPQVYTRLPFLFRDWIKENTG 240  
Db 795 SGGVDSQCGSGGGLSSVEADGRIFQAGVYSGDGCQAQRNKPQVYTRLPFLFRDWIKENTG 854  
QY 241 V 241  
Db 855 V 855

RESULT 2

*Comparison*

```

US-08-944-483-65
; Sequence 65, Application US/08944483
; Patent No. 6232456
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KLASS, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STROUPE, STEVEN D.
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
; TITLE OF INVENTION: OF THE PROSTATE
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,483
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6183.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 235 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6232456e
; US-08-944-483-65

```

[illegible]

```

1  RESULT      3
2  US-08-807-151-3
3  ; Sequence 3, Application US/08807151
4  ; Patent No. 6043033
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Bandman, Olga
7  ; APPLICANT: Lai, Preeti
8  ; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
9  ; TITLE OF INVENTION: PROTEASE
10 ; NUMBER OF SEQUENCES: 5
11 ; CORRESPONDENCE ADDRESS:
12 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
13 ; STREET: 3174 Porter Drive
14 ; CITY: Palo Alto
15 ; STATE: CA
16 ; COUNTRY: US
17 ; ZIP: 94304
18 ; COMPUTER READABLE FORM:
19 ; MEDIUM TYPE: Diskette
20 ; COMPUTER: IBM Compatible
21 ; OPERATING SYSTEM: DOS
22 ; SOFTWARE: FastSEQ Version 2.0
23 ; CURRENT APPLICATION DATA:
24 ; APPLICATION NUMBER: US/08/807,151
25 ; FILING DATE: Filed Herewith
26 ; CLASSIFICATION: 424
27 ; PRIOR APPLICATION DATA:
28 ; APPLICATION NUMBER:
29 ; FILING DATE:
30 ; ATTORNEY/AGENT INFORMATION:
31 ; NAME: Billings, Lucy J.
32 ; REGISTRATION NUMBER: 36,749
33 ; REFERENCE/DOCKET NUMBER: PF-0227 US
34 ; TELECOMMUNICATION INFORMATION:
35 ; TELEPHONE: 415-855-0555
36 ; TELEFAX: 415-845-4166
37 ; INFORMATION FOR SEQ ID NO: 3:
38 ; SEQUENCE CHARACTERISTICS:
39 ; LENGTH: 235 amino acids
40 ; TYPE: amino acid
41 ; STRANDEDNESS: single
42 ; TOPOLOGY: linear
43 ; IMMEDIATE SOURCE:
44 ; LIBRARY: GenBank
45 ; CLONE: 416132
46 ;
47 ; US-08-807-151-3

```

```

Query Match      41.3%; Score 544.5; DB 3; Length 235;
Best Local Similarity 43.5%; Pred. No. 1.2e-49;
Matches 103; Conservative 50; Mismatches 77; Indels 7; Gaps 4

```

Qy	1	VVGGTDADEGEWPNVSLHALGOGHTCGASLIISPNWLVSAAHCYIDDRGFYSDPQTQWTA	60
Dd	1	IVGGSDREGAWPVVALFDQQ-VQCASLSVRDLVSAAHCVYG---RNMEPSKWA	55
Qy	61	FLGLHDSQRSAPGVGERLKRIRISHPFNDFTDYDIALLELEKKPEYSSMVRPCLPD	120
Dd	56	VLGLHMASNLTSPOIELRIDQIVINPHYNRKRKNNDIAMHLEMKMYNTDYIQPICLPE	115
Qy	121	ASHVFPAKAIWTGHGHGYOYGGTGAIILOKEIRVINQTTCCENLLPQ-OITPRMCVGF	179
Dd	116	ENQVFPPRICSTAGWALTYLGSTDADVLQEADVPLLNKCKOQMFEYNITENVCAGY	175
Qy	180	LSGGVDSQCQDSGGPISSSVEADGRIFQAGVSWGDGCQAQRNKPGVYRLPLFRDWIK	236
Dd	176	EAGGVDSQCQDSGGPLWCQE-NNRNLLAGVTSGFYQCQCALPNRPGVYARVPRTFETWIQ	231

```

RESULT          4
US-9478-957-3
: sequence 3, Application US/09478957

```

```

1  NUMBER OF SEQUENCES: 38
2  CORRESPONDENCE ADDRESS:
3  ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
4  STREET: 87 CambridgePark Drive
5  CITY: Cambridge
6  STATE: MA
7  COUNTRY: USA
8  ZIP: 02140
9  COMPUTER READABLE FORM:
10 MEDIUM TYPE: Floppy disk
11 COMPUTER: IBM PC compatible
12 OPERATING SYSTEM: PC-DOS/MS-DOS
13 SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
14 CURRENT APPLICATION DATA:
15 APPLICATION NUMBER: US/08/200,900A
16 FILING DATE: 23-FEB-1994
17 CLASSIFICATION: 435
18 ATTORNEY/AGENT INFORMATION:
19 NAME: Melnert, Maureen C.
20 REGISTRATION NUMBER: 31,544
21 REFERENCE/DOCKET NUMBER: GI 5201-FWC
22 TELECOMMUNICATION INFORMATION:
23 TELEPHONE: (617) 876-1170 X8574
24 TELEFAX: (617) 876-5851
25 INFORMATION FOR SEQ ID NO: 2:
26 SEQUENCE CHARACTERISTICS:
27 LENGTH: 798 amino acids
28 TYPE: amino acid
29 STRANDEDNESS: single
30 TOPOLOGY: linear
31 MOLECULE TYPE: protein
32 US-08-200-900A-2
33
34 Query Match 41.3%; Score 544.5; DB 1; Length 798;
35 Best Local Similarity 43.5%; Pred. No. 6.le-49;
36 Matches 103; Conservative 50; Mismatches 77; Indels 7; Gaps
37
38 QY 1 VVGGTDADEGWPQVSLHALGOGHCIGASLISPNKLYSAAHCYIDDRGFYSDPTQWTA 60
39 :|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
40 Db 564 IVGGSDSREGAPWVVALYFDQ-QVCGASLSVRDLVLSAAHCYVG----RNNEPSKWA 618
41 :|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
42 QY 61 FLGLHDSQRSAPGVQERRUKRIISHFFNDFDYYDIALLELEKPAEYSSMYRPICLPD 120
43 :|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
44 Db 619 VLGLHMASNLTSQIETRLIDQIVINPHYNKRRKKNDIAMHLEMKVNYTDYIQICLPE 678
45
46 QY 121 ASHVPFPAKALWTGHWGHTQYGGTGALILKEGEIRVINQTCENLLPQ-QITPRMVCVGF 179
47 :|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
48 Db 679 ENQVFPFGRICISAGWALLYQGSTADVLQEDADVPLLSNEKCOQMPEYNTENMYCAGY 738
49
50 QY 180 LSGVYDCQDGGSLSSVBEADQIFPAQVSWGSDGCAQRNKPQVYTRPLPLFRDWIK 236
51 :|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
52 Db 739 EAGGVDSQCQDGGSLPQCE-NNRWLLAGVTSFGYQCALPNRPQGVYARVPFTEWIQ 794
53
54 RESULT 6
55 PCT-US94-00616-2
56 ; Sequence 2, Application PC/TUS9400616
57 ; GENERAL INFORMATION:
58 ; APPLICANT:
59 ; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
60 ; NUMBER OF SEQUENCES: 33
61 ; COMPUTER READABLE FORM:
62 MEDIUM TYPE: Floppy disk
63 COMPUTER: IBM PC compatible
64 OPERATING SYSTEM: PC-DOS/MS-DOS
65 SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
66 CURRENT APPLICATION DATA:
67 APPLICATION NUMBER: PCT/US94/00616
68 FILING DATE:
69 CLASSIFICATION:
70 INFORMATION FOR SEQ ID NO: 2:
71 SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-00616-2

Query Match 41.3%; Score 544.5; DB 5; Length 798;
Best Local Similarity 43.5%; Pred. No. 6.1e-49;
Matches 103; Conservative 50; Mismatches 77; Indels 7; Gaps 4;

QY 1 VVGTTDADGEWPMQVSLHALGQGHICGASLISPNMLVSAAHYIDDRGFRYSDDPTQWTA 60
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 564 IVGSDSREGAPWVVALYFDQ-QVCGASLVSRDLVSAAHCVG---RNMEPSKWA 618
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 61 FLGLHQSORAPGVQERLRIISHPFNFDTFDYDIALLELEKPAEYSSMVRPCLPD 120
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 619 VLGLHMASNLTPQIETRLIDQIVINPHYNKRKNND:AMHLEKNYNTYIQPICLPE 678
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 121 ASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINOTTCEMLLPQ-QITPRMVCVF 179
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 679 ENQVFPGRICSTAGWALIYGGSTADVLQEADVPLLSNEKQQCMPEYNTENMWCAGY 738
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 180 LSGVDSCQDGGSGPLSSVEADGRIFQAGVYSWGDCGAQRNKPQGYTRPLPLFRDWIK 236
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 739 EAGVDSCQDGGSGPLMCOE-NNRWLLAGVTSFGYQCALPNRPQGYAKVRPTWMIQ 794
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 7
US-08-508-448C-19
; Sequence 19, Application US/08508448C
; Patent No. 5804410
; GENERAL INFORMATION:
; APPLICANT: Kazuyoshi YAMAOKA et al.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCE ENCODING
; TITLE OF INVENTION: TRYPSIN-LIKE ENZYME AND PROCESS FOR PRODUCING THE ENZYME
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/508,448C
; FILING DATE: July 28, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 232 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-508-448C-19

```



RESULT 10  
3-08-807-151-1  
Sequence 1, Application US/0807151  
Patent No. 6043033  
GENERAL INFORMATION:

```

RESULT 11
US-09-478-957-1
; Sequence 1, Application US/09478957
; Patent No. 6350448
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
; TITLE OF INVENTION: PROTEASE

```

NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: US  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/478,957  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/807,151  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0227 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 283 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: SCORNO01  
CLONE: 556016  
US-09-478-957-1

Query Match 37.0%; Score 488; DB 4; Length 283;  
Best Local Similarity 41.0%; Pred. No. 1.4e-43;  
Matches 98; Conservative 45; Mismatches 82; Indels 14; Gaps 7;  
QY 1 VVGTTDADGEMPMQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFRYSPTQWTA 60  
DB 47 IVGGSALPGAPMQVSLH-VQNVHVCGGSIITPEWIVTAHCKEKP-...LNPHWHTA 101  
QY 61 FLGLHDQS-QRSAPGVQERRKRIISHPFNDFTDYDIALLELEKPAEYSSMVRPICLP 119  
DB 102 FAGILLRQSPMFYAGYQ---VEKVISHPNYDSKTKNNNDIALMKLQKPLTFNDLVKPYCLP 158  
QY 120 DASHVEPPAKATWVTGWGHTQYGGTGALILQKGEIRVINQTTCTEN--LLPQQTTPRMVCV 177  
DB 159 NFGMMLQPEQLCWSGWGATEKGTSEVLNAKVLLIETQRCNSRVYVNDLITPAMICA 218  
QY 178 FLGSGVDSCQDGGPLSSVEADGRI--FOAGVSVWSDGCAQRNKPQVYTRLPFLFRDWI 235  
DB 219 GFLOQNVDSQDGGSL--VTSKNNIWWLIGDTSWGGSCAKAYRPGVYGVNMVFTDWI 275  
RESULT 12  
US-09-518-046-2  
Sequence 2, Application US/09518046  
Patent No. 6294663  
GENERAL INFORMATION:  
APPLICANT: O'Brien, Timothy J.  
APPLICANT: Underwood, Lowell J.  
TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed  
TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof  
FILE REFERENCE: D6192CIP  
CURRENT APPLICATION NUMBER: US/09/518,046  
CURRENT FILING DATE: 2000-03-02  
EARLIER APPLICATION NUMBER: 09/261,416

EARLIER FILING DATE: 1999-03-03  
NUMBER OF SEQ ID NOS: 153  
SEQ ID NO 2  
LENGTH: 454  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: complete amino acid sequence of TADG-12  
US-09-518-046-2  
Query Match 36.8%; Score 485; DB 4; Length 454;  
Best Local Similarity 40.4%; Pred. No. 5.3e-43;  
Matches 97; Conservative 44; Mismatches 85; Indels 14; Gaps 6;  
QY 1 VVGTTDADGEMPMQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFRYSPTQWTA 60  
DB 218 IVGGMNLSLLSQMPQASLQFGY-HLCGGSVITPLWITTAHCVYD----LYLPKSWTI 271  
QY 61 FLGLHDQS-QRSAPGVQERRKRIISHPFNDFTDYDIALLELEKPAEYSSMVRPICLP 120  
DB 272 QVGLVSLLDNPAP---SHLVEKIVYHSKYKPKRLGNDIALMKLAGPLTFNEMIQVCLPN 328  
QY 121 ASHVEPPAKATWVTGWGHTQYGGTGALILQKGEIRVINQTTCTEN--ENLLPQQTTPRMVCV 178  
DB 329 SEENFPDGKVCWTSGWGATEDGGDASPVLNHAAPLISNKICNHRDVGIGIISPSMLCAG 388  
QY 179 FLGSGVDSCQDGGPLSSVEADGRIQF-AGVSVWSDGCAQRNKPQVYTRLPFLFRDWI 237  
DB 389 YLTGGVDSCQDGGSL--VCQERRLMKLVGATSGFSGCAEVNKNKPGVYTRVTSFLOWIHE 446  
RESULT 13  
US-09-000-846-2  
Sequence 2, Application US/09000846  
Patent No. 5981830  
GENERAL INFORMATION:  
APPLICANT: WT, QINGYU  
APPLICANT: SADLER, JASPER  
TITLE OF INVENTION: KNOCKOUT MICE AND THEIR PROGENY WITH  
TITLE OF INVENTION: DISRUPTED SERINE PROTEASE GENES  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MILLEN, WHITE, ZELANO & BRANIGAN, P.C.  
STREET: 2200 CLARENDON BLVD. SUITE 1400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: US  
ZIP: 22201  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/000,846  
FILING DATE: 30-DEC-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/866,058  
FILING DATE: 30-MAY-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: LEBOVITZ, RICHARD M.  
REGISTRATION NUMBER: 37,067  
REFERENCE/DOCKET NUMBER: BERLX 65P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-243-6333  
TELEFAX: 703-243-6410  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 416 amino acids  
TYPE: amino acid

```

;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
US-09-000-846-2

```

Query Match	36.5%	Score 481.5;	DB 2;	Length 416;
Best Local Similarity	40.1%;	Pred. No. 1.1e-42;		
Matches 99; Conservative	41;	Mismatches 86;	Indels 21;	Gaps 8;

[illegible]

RESULT 14  
US-08-944-483-67  
; Sequence 67, Application US/06944483  
; Patent No. 6232456  
; GENERAL INFORMATION:  
; APPLICANT: COHEN, MAURICE  
; APPLICANT: COLPITTS, TRACEY L.  
; APPLICANT: FRIEDMAN, PAULA N.  
; APPLICANT: GRANADOS, EDWARD N.  
; APPLICANT: KLASS, MICHAEL R.  
; APPLICANT: RUSSELL, JOHN C.  
; APPLICANT: STEWART, KENT D.  
; APPLICANT: STROUPE, STEVEN D.  
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS  
; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES  
; TITLE OF INVENTION: OF THE PROSTATE  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60054-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/944,483  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Becker, Cheryl L.  
; REGISTRATION NUMBER: 35,441  
; REFERENCE/DOCKET NUMBER: 6183.US.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 847/935-1729  
; TELEFAX: 847/938-3623

TELEX: 67;  
 INFORMATION FOR SEQ ID NO: 67;  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 255 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: No. G232456  
 PS-08-04-483-67

```

Query Match      35.9%; Score 473.5; DB 4; Length 255;
Best Local Similarity 40.7%; Prad. No. 4.le-4.2;
Matches 100; Conservative 38; Mismatches 89; Indels 19; Gaps
QY 1 VYGGTDADGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYIDORGFRYSDPTQWTA 60
   :||| :||| ||||| | :||| :||| :||| :||| :||| :||| :||| :|||
Db 1 IVGGRTDSLGRWPQVSLRYDG-AHLGGGSLSGDWLTAACHPE-----RNRVLSRWRY 55
QY 61 FLGLHDQSRGAPGVQZERLKRRIISH-----PFF--NDTFFDYDIALLELEKPAEYSMWY 114
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 56 FAGA--VAQASPHGLQ-LGVQAVYVYGGVLPFRDPNSENSNDIALVHLSPLLEYIQ 112
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 115 P1CLPDASHEVPFAGKAIWVGWGTGYGGTGALILQKEIRVINQTTCE--NLLPQOITP 172
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 113 PVELPAAGALVDGKICTVGTGWGNTQYVYQQAGVLQEARVPIISNDVCGADFVYGNQIKP 172
QY 173 RMCMCVFGLSGVDSCQDSSGGLP---SSVEADGRIFQAGVYVSWDGGCAQRNKGQVYTRLP 229
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 173 KMFCAGYPEGGIDACQDSSGPFVCEDSISRTFRNLCKIVSWGTGCGALACKPGVYTKVS 232
QY 230 LFRDWI 235
   ||| :|||
Db 233 DFERWI 238

```

```

RESULT 15
US-09-027-337-3
; Sequence 3, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotochi
; TITLE OF INVENTION: TAGD-15: An Extracellular Serine Protease Overexpressed in
; TITLE OF INVENTION: Breast and Ovarian Carcinomas
; FILE REFERENCE: D6064
; CURRENT APPLICATION NUMBER: US/09/027.337B
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO. 3
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Serine protease catalytic domain of hepsin (Heps)
; OTHER INFORMATION: homologous to similar domain in TAGD-15
; US-09-027-337-3

```

```

Query Match      35.9%; Score 473.5; DB 2; Length 256;
Best Local Similarity 40.7%; Pred. No. 4.1e-42;
Matches 100; Conservative 38; Mismatches 89; Indels 19; Gaps 8

QY 1 VVGGTADAGEPWPQVSLHALGQGHICGASISFNWLVSAAHCYIDRGFRYSDPTQWTA 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2 IVGGRDTSLGRWPQVQVSLRYDYG-AHLGCGSLSGDWMVLTAAHCFFE---RNVLVSRWRV 56
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 61 FLGLHQDSQSPAGVQVRLKRLISH---PFF--NDFTDYDFALLELEKPAEYSSMYR 114
   | | | | | : : : : : | | | | | : : : : : : : : : : : : : : : :
Db 57 FAGA--VAQASPHGQLQ-LGVQAVVTHGGLYLFPRDPNSEENSNDIALVHLSSPLPLEYIQ 113
   | | | | | : : : : : | | | | | : : : : : : : : : : : : : : : :

QY 115 PICLPDASHVFFPAGKAIWTVTGMGHGTQYGGTGALILQKGEIRVNTQTTCF--NLLPQOITP 172
   | : | : | | | | | : : : : : | : : : : : | : : : : : | : : : : |

```

Db 114 PVCLPAAGQALVDGKICTVTGNGNTQYGGQAGVLQEARVPIISNDVCNGADFYGNQIKP 173  
QY 173 RMCVGFSLGGVDSOGDSGGPL---SSVEADGRIFQAGVWSWGDGCAORNKPGVYTRLP 229  
Db 174 KMFCAGYPEGGIDACOGDSGGPFVCEDSISRTPRWRLCGIVSWGTGCALAOXKPGVYTKVS 233  
QY 230 LFRDWI 235  
Db 234 DFREWI 239

Search completed: July 22, 2002, 17:12:25  
Job time: 226 sec